Spatio-temporal variability of temperature effects on mortality in São Paulo, Brazil

Aina Roca-Barceló

Department of Epidemiology and Biostatistics

School of Public Health

Faculty of Medicine

Imperial College London

Submitted in part fulfilment of the requirements for the degree of

Doctor of Philosophy of Imperial College London

August 2023

Statement of Originality

I certify that the work presented here is solely my own unless otherwise explicitly stated, cited or referenced. I have benefited from the support, advice and guidance of my supervisors, collaborators and assessors, whose contributions I acknowledge in the Acknowledgements.

Copyright Statement

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution Non Commercial 4.0 International Licence (CC BY-NC). Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe the changes. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Abstract

Non-optimal temperatures pose significant health risks, yet understanding the nuanced relationship between temperature and mortality within cities, including vulnerability among specific populations and neighbourhoods. The lack of temperature data at the necessary spatial and temporal resolution, albeit not the only cause, an important one. Addressing this research gap is crucial for identifying shortcomings in current strategies and prioritizing interventions for high-risk groups/areas. My thesis focuses on São Paulo to characterize temperature-related all-cause mortality risk over time and across population and area-level characteristics.

To analyse adaptation patterns, I examined temperature-mortality trends in São Paulo State from 2000 to 2018. Using Time-Varying Distributed Lag Non-linear Models (tv-DLNM), I captured the non-linear and delayed effects of temperature on mortality, adjusting for various factors and stratifying the analysis by gender, age group, and ethnicity. Significant trends in Minimum Mortality Temperature (MMT) and Cumulative Relative Risk (cRR) were found, with variations across population groups indicating disparities in adaptation capacities.

Spatial variability in temperature across São Paulo was assessed using a Random Forest (RF) model, predicting daily mean temperature at 500m2 resolution. The model integrated data from ground measurement stations and satellite-derived products, exhibiting good overall performance and outperforming linear models. With these temperature estimates and individual mortality records, I explored differences in temperature-mortality associations among population groups based on individual and area-level factors. Findings revealed elevated mortality risks among older individuals, women, and those with limited education for both heat and cold, and among married or widowed individuals and those of mixed ethnic backgrounds for cold temperatures. Results for contextual factors were inconsistent, warranting further research.

This thesis delves into São Paulo's temperature-mortality link, stressing the need for continuous public health monitoring and updates. The development of a spatiotemporally resolved temperature dataset, first-of-its-kind in Latin America, minimizes exposure bias and facilitates small-area studies.

Contents

List of Tables **8**

List of Figures **10**

List of Abbreviations **15**

Acknowledgments **16**

1 Introduction **18**
1.1 Rationale . . . 18
1.2 Aims and objectives . . . 19
1.3 Structure of the thesis . . . 20
1.4 Ethics and governance . . . 20
1.5 Data sharing . . . 20

2 Background **21**
2.1 Disaster risk: the case of non-optimal temperatures . . . 21
2.1.1 A dynamic and heterogeneous phenomenon . . . 23
2.2 Impact: the health burden of non-optimal temperature . . . 24
2.2.1 Physiological pathways . . . 24
2.2.2 Temperature measures . . . 26
2.2.3 Temperature-mortality relationship . . . 28
2.2.4 Study designs: time-series and case-crossover . . . 29
2.2.5 Accounting for non-linear and delayed effects: the DLNM framework . . . 30
2.2.6 Temporal variability in temperature-related health impacts . . . 32
2.2.7 Spatial variability in temperature-related health impacts . . . 32
2.3 Hazard: ambient temperature . . . 34
2.3.1 Overview . . . 34
2.3.2 Temperature data types . . . 35
2.3.2.1 In situ recordings . . . 36
2.3.2.2 Climate models and their downscaled datasets . . . 36
2.3.2.3 Reanalysis data . . . 37
2.3.2.4 Interpolation methods . . . 38
2.4 Vulnerability and exposure: determinants . . . 44
2.4.0.1 Individual determinants . . . 45
2.4.0.2 Contextual determinants . . . 46
2.5 Scale matters: the special case of urban settings . . . 48
2.5.1 City-level studies . . . 48
2.5.2 Intra-city studies . . . 49
2.6 Public health strategies and temperature . . . 51
2.6.1 Health and climate change plans or strategies . . . 52
2.6.2 Climate-informed health surveillance and early warning systems . . . 52
2.7 São Paulo, Brazil . . . 54
2.7.1 Temperature: past, present and future . . . 55

2.7.2 Individual and contextual determinants . . . 59
2.7.3 Climate adaptation policy in São Paulo . . . 59
2.7.4 Heath and heat strategies . . . 60
2.7.5 Temperature-mortality epidemiological evidence . . . 61
2.8 Summary . . . 64

3 Exploring disparities in adaptation to cold and heat 65
3.1 Overview and objectives . . . 65
3.2 Data and Methods . . . 66
3.2.1 Study area . . . 66
3.2.2 Data sources . . . 67
3.2.2.1 Mortality data . . . 67
3.2.2.2 Temperature data . . . 67
3.2.2.3 Confounder data . . . 68
3.2.2.4 Vulnerability data . . . 72
3.2.3 Statistical analyses . . . 74
3.2.3.1 Trends in daily mean temperature . . . 74
3.2.3.2 tv-DLNM: overview of the modelling framework . . . 75
3.2.3.3 Indicators of adaptation . . . 76
3.2.3.4 Investigating if the observed trends are associated to changes in climate 77
3.2.4 Sensitivity analyses . . . 78
3.2.4.1 Model parametrization . . . 78
3.2.4.2 Relative humidity and air pollution . . . 78
3.2.4.3 Temperature gridded datasets . . . 81
3.2.5 Software . . . 81
3.2.6 Data access disclaimer . . . 82
3.3 Results . . . 82
3.3.1 Descriptive statistics . . . 82
3.3.1.1 Temporal trends in meteorological and air pollution data . . . 82
3.3.1.2 Temporal trends in mortality data . . . 84
3.3.2 Disparities in adaptation . . . 86
3.3.2.1 Minimum mortality temperature (MMT) . . . 86
3.3.2.2 Cumulative relative risk (cRR) . . . 89
3.3.2.3 Is climate driving the changes? . . . 95
3.3.3 Sensitivity analyses . . . 99
3.3.3.1 Model parametrization . . . 99
3.3.3.2 Relative humidity and air pollution . . . 99
3.3.3.3 Gridded dataset . . . 99
3.4 Discussion . . . 101
3.4.1 Temperature-mortality association in São Paulo . . . 101
3.4.2 Trends in temperature-mortality association: adaptation . . . 102
3.4.3 Temperature and confounder data . . . 104
3.4.4 Public health implications . . . 106
3.4.5 Strengths and novelty . . . 106
3.4.6 Limitations . . . 106
3.5 Summary . . . 108

4 Statistical modelling of daily mean temperature at high spatial resolution 110
4.1 Overview and objectives . . . 110
4.2 Data and Methods . . . 111
4.2.1 Rationale for the Random Forest . . . 111
4.2.2 Study area . . . 112
4.2.3 Data sources . . . 114

4.2.3.1 Temperature data . . . 114
4.2.3.2 Input features . . . 114
4.2.4 Pre-modelling data wrangling . . . 124
4.2.4.1 Weather stations data processing . . . 124
4.2.4.2 Input feature data processing . . . 134
4.2.4.3 Creating and preparing input dataset . . . 140
4.2.4.4 Training and external validation dataset . . . 140
4.2.5 Model building . . . 142
4.2.5.1 Random Forest (RF) . . . 142
4.2.5.2 Hyper-parameter tuning and feature selection . . . 143
4.2.5.3 Feature importance . . . 145
4.2.6 Model validation . . . 146
4.2.6.1 Spatial cross-validation . . . 146
4.2.6.2 External validation . . . 146
4.2.6.3 Metrics for evaluating model accuracy . . . 147
4.2.7 Model comparison . . . 148
4.2.7.1 Land use regression, LUR . . . 148
4.2.7.2 Prediction and performance comparison . . . 149
4.2.8 Sensitivity analyses . . . 149
4.2.9 Data access disclaimer . . . 149
4.3 Results . . . 150
4.3.1 Descriptive statistics of the response variable and input features . . . 150
4.3.2 Random Forest outputs . . . 157
4.3.2.1 Model training . . . 157
4.3.2.2 Feature importance . . . 159
4.3.2.3 Model predictions . . . 160
4.3.2.4 Model validation . . . 165
4.3.3 Model comparison: LUR . . . 174
4.3.4 Sensitivity analyses . . . 181
4.4 Discussion . . . 183
4.4.1 Strengths . . . 191
4.4.2 Limitations . . . 192
4.5 Summary . . . 193

5 Spatio-temporal variability in the temperature-mortality association 194
5.1 Overview and objectives . . . 194
5.2 Data and Methods . . . 196
5.2.1 Study design . . . 196
5.2.2 Data sources . . . 196
5.2.2.1 Temperature data . . . 196
5.2.2.2 Mortality data . . . 197
5.2.2.3 Confounder data . . . 198
5.2.2.4 Individual determinants . . . 198
5.2.2.5 Contextual determinants . . . 199
5.2.3 Data linkage . . . 206
5.2.4 Statistical Analyses . . . 207
5.2.4.1 Overview of the modelling framework and model specification . . . 207
5.2.4.2 Stratified analyses by individual and contextual determinants . . . 209
5.2.4.3 Sensitivity analyses . . . 209
5.2.5 Software . . . 209
5.2.6 Data access disclaimer . . . 209
5.3 Results . . . 212
5.3.1 Descriptive statistics . . . 212

5.3.2 Temperature-mortality association . . . 213
5.3.3 Individual determinants . . . 215
5.3.4 Contextual determinants . . . 217
5.4 Discussion . . . 221
5.4.1 Overall temperature-mortality association in São Paulo . . . 221
5.4.2 Individual determinants . . . 222
5.4.3 Contextual determinants . . . 224
5.4.4 Counter-intuitive results and comparability considerations . . . 229
5.4.5 Strengths and novelty . . . 231
5.4.6 Limitations . . . 232
5.5 Summary . . . 233

6 Discussion **235**
6.1 Research in context . . . 235
6.1.1 Evidence before this study . . . 235
6.1.2 Added value of this thesis . . . 237
6.2 Strengths and novelty . . . 239
6.3 Limitations . . . 240
6.4 Public health and policy implications . . . 240
6.4.1 Short-term interventions . . . 241
6.4.2 Long-term interventions . . . 242
6.5 Future work . . . 243
6.5.1 Temperature modelling . . . 244
6.5.1.1 External validation . . . 244
6.5.1.2 Improved input data . . . 244
6.5.1.3 Modelling outside of the training space . . . 245
6.5.1.4 Model Transferability . . . 247
6.5.1.5 Operational EO for heat monitoring . . . 249
6.5.1.6 Interdisciplinary collaboration . . . 250
6.5.2 Quantification of the temperature-attributable health burden . . . 250
6.5.2.1 Untangling the mechanism behind adaptation . . . 250
6.5.2.2 Need for collaborative, systems-thinking and target-driven research . . . 251
6.5.2.3 Exploring spatiotemporal patterns . . . 251
6.5.2.4 Detailed investigation of the temperature-related health burden . . . 252
6.5.2.5 In-hospital mortality . . . 253
6.5.2.6 Artificial neural networks . . . 254
6.6 Conclusions . . . 254

Bibliography **256**

Appendix **296**
6.7 Information . . . 296
6.7.1 Data sharing . . . 296
6.7.2 Collaboration with the University of São Paulo . . . 298
6.7.3 Research visit at the European Space Agency ϕ-lab . . . 298
6.8 Tables . . . 301
6.9 Figures . . . 304

List of Tables

2.1 Epidemiological studies on health impacts of temperature in São Paulo 63

3.1 Description of air pollution network spatial representativeness 70
3.2 Missing data per population group 73
3.3 Missing data by year for variable "years of education" 73
3.4 Sensitivity analyses 79
3.5 Summary statistics meteorological and air pollution data 82
3.6 Summary statistics mortality data 86
3.7 MMT for 2000 and 2018 by population group 87
3.8 cRR for 2000 and 2018 by population group 92
3.9 ERA5-Land and IAG-USP temperature-mortality association: cRR and MMT 101

4.1 Candidate input features 123
4.2 LST imputation performance metrics 137
4.3 Stations in training and external validation sets 141
4.4 Hyper-parameter definitions. 143
4.5 Sensitivity analyses 149
4.6 Descriptive statistics 153
4.7 RF and LUR model accuracy 171
4.8 Model accuracy by year 171
4.9 External validation performance scores 172
4.10 Model performance and accuracy by sensitivity analysis 181

5.1 Annual success rate of mortality data geo-coding 198
5.2 Descriptive statistics by HDUs, São Paulo Metropolitan area 201
5.3 Contextual determinants selected, codes and description 202
5.3 Contextual determinants selected, codes and description 203
5.4 Code and description of the macroareas 206
5.5 Model parametrization tuning 211
5.6 Temperature summary statistics 212

5.7 cOR by individual determinant categories 216
5.8 cOR by contextual determinant categories 218
5.8 Cumulative Odds Ratio (cOR) by contextual determinant categories 219

6.1 Individual determinants re-coded 301
6.2 Temperature statistics by contextual determinant 302
6.3 Temperature statistics by individual determinant 303

List of Figures

2.1 Disaster risk conceptual framework 22
2.2 Time-stratified case-control matching 30
2.3 Examples of visual representations of DLNM outputs 31
2.4 Historical spatial and temporal variability in annual excess deaths for cold and heat . 34
2.5 Scheme of a Random Forest 44
2.6 Map of São Paulo 55
2.7 Maximum and minimum temperature in São Paulo 56
2.8 Annual mean temperature for São Paulo between 1980 to 2020 56
2.9 Temperature projections for the State of São Paulo under RCP4.5 and 8.5 58

3.1 Location of IAG-USP meteorological station 68
3.2 Air pollution monitoring station network 69
3.3 Pairwise correlation of PM_{10} estimates across stations 71
3.4 Diagram of adaptation indicators 77
3.5 Meteorological and air pollution daily and annual estimates 83
3.6 ENSO-patterns and IAG-USP temperature data 84
3.7 Time series of death counts 85
3.8 Annual MMT estimates by population group 88
3.9 Annual cRR by population group 90
3.10 Annual cumulative relative risk (cRR) by population group 93
3.11 Visual representation of the interaction terms by population group 94
3.12 Association between MMT and annual mean temperature 96
3.13 Association between cRR for extreme cold and annual extreme cold 97
3.14 Association between cRR for extreme heat and annual extreme heat 98
3.15 ERA5-Land and IAG-USP temperature comparisons 100
3.16 ERA5-Land and IAG-USP temperature-mortality association 101

4.1 Study area and location of meteorological stations 113
4.2 River index variable 119

4.3 Heatmap of annual data completeness by station . . . 124
4.4 Meteorological station spatial clusters . . . 125
4.5 Meteorological stations duplicates . . . 126
4.6 Histogram and normal Q-Q plot of temperature by cluster . . . 127
4.7 Box plot of station-specific temperature by cluster . . . 128
4.8 Station-specific missing values by year . . . 130
4.9 Station-specific missing values by month and day of the week . . . 131
4.10 Missing temperature by network and cluster . . . 132
4.11 Missing temperature by SES and rurality . . . 133
4.12 Workflow of the imputation and spatiotemporal harmonization of features . . . 134
4.13 Gapfill prediction window hyper-parameter tuning . . . 139
4.14 Example of rasters imputed with *gapfill* . . . 139
4.15 Illustration of data split: train, test, and external validation . . . 141
4.16 Density plot of daily temperature for hyperparameter tuning . . . 144
4.17 Temperature from meteorological stations, spatial and temporal variability . . . 151
4.18 Maps of the input features . . . 152
4.19 Pearson correlation matrix input features . . . 154
4.20 Pairwise Pearson correlation with temperature . . . 155
4.21 Density scatter plot between daily temperature and input features . . . 156
4.22 Model workflow . . . 157
4.23 Hyper-parameter tuning . . . 158
4.24 Heatmap selected features per hyper-parameter combination . . . 159
4.25 RF feature importance . . . 160
4.26 Map of temperature predictions for all period and random day . . . 161
4.27 Spatial distribution of predicted temperatures . . . 162
4.28 RF temperature predictions by year . . . 163
4.29 RF temperature predictions by month . . . 164
4.30 Density scatter observed vs. RF predicted - all period . . . 166
4.31 Difference between observed and predicted temperature . . . 166
4.32 Annual differences in model accuracy and $\Delta Temp$ . . . 167
4.33 Monthly differences in model accuracy and $\Delta Temp$ . . . 169
4.34 RMSE per station for the period 2015-2019 . . . 170
4.35 Rural/urban differences in model accuracy and $\Delta Temp$ . . . 170
4.36 Time series and scatter plot of RF external validation . . . 173
4.37 LUR model residual diagnostic plots . . . 174

4.38 LUR regression coefficients . . . 175
4.39 Map of the average daily mean temperature predicted by RF and LUR for all period . 176
4.40 Map of the average daily mean temperature predicted by RF and LUR, by year 178
4.41 Boxplot of the average temperature predicted by RF and LUR overall, by year and month 179
4.42 Scatter plots of LUR spatial CV and external validation . . . 180
4.43 Map of predicted temperature overall, by sensitivity analyses . . . 182

5.1 Administrative boundaries São Paulo . . . 201
5.2 IDHM 2010 in the São Paulo Metropolitan Area . . . 204
5.3 Urban planning variables . . . 206
5.4 Data linkage overview . . . 207
5.5 Correlation plot for contextual determinants . . . 213
5.6 cOR of all non-external cause mortality associated to daily mean temperature 214
5.7 Sensitivity analyses results . . . 215
5.8 cOR by individual determinant categories . . . 217

6.1 Project's website landing page . . . 297
6.2 Cumulative temperature-mortality association for 2000-18 by population group 304
6.3 Temperature-mortality association by education group . . . 305
6.4 Trends of MMT by sensitivity analyses . . . 306
6.5 Trends of the cRR for extreme heat by sensitivity analysis . . . 307
6.6 Trends of the cRR for extreme cold by sensitivity analysis . . . 308
6.7 Missing values . . . 309
6.8 Daily and weekly and temperature from included meteorological stations . . . 309
6.9 Feature selection agreement by hyper-parameter tuning run . . . 310
6.10 Observed and predicted temperature at each station in 2015 . . . 311
6.11 RMSE per station for each year . . . 312
6.12 Density scatter plot of RF and LUR predicted temperatures overall and by year . . . 313
6.13 Annual differences in model accuracy and $\Delta Temp$ for LUR model . . . 314
6.14 Monthly differences in model accuracy and $\Delta Temp$ for LUR model . . . 315
6.15 RMSE per station for the period 2015-2019 . . . 316
6.16 Rural/urban differences in model accuracy and $\Delta Temp$ for LUR model . . . 316
6.17 Correlation plot between temperature and NDVI scores . . . 317
6.18 Histogram of population and sample contextual determinants (continuous) . . . 318
6.19 cOR associated to heat by contextual determinant quintiles . . . 319
6.20 cOR associated to cold by contextual determinant quintiles . . . 320

List of Abbreviations

AIC	Akaike Information Criterion.
AMT	Annual Mean Temperature.
AOA	Area of Applicability.
bsa	Black Sky Albedo.
CEP	*Código de Endereçamento Postal*, equivalent to postal code.
CETESB	*Companhia Ambiental do Estado de São Paulo*.
CI	Confidence Intervals.
cOR	Cumulative Odds Ratio.
cRR	Cummulative Relative Risk.
CV	Cross-validation.
CVD	Cardiovascular Diseases.
df	Degrees of Freedom.
DLNM	Distributed Lag Non-linear Models.
ECMWF	European Centre for Medium-Range Weather Forecasts.
ENSO	El Niño Southern Oscillation.
ERA5	ECMWF Re-Analysis 5^{th} generation.
ESA	European Space Agency.
FFS	Forward Feature Selection.
GDP	Gross Domestic Product.
GEE	Google Earth Engine.
HDU	Human Development Units.
HHAP	Heat-Health Action Plans.
HHWS	Heat Health Warning System.
HW	Heat wave.
IAG-USP	*Instituto Astronomia, Geofísica e Ciencias Atmosféricas, Universidade de São Paulo*.
ICD-10	International Classification of Disease 10^{th} edition.

IDHM Human Development Index.

IDW Inverse Distance Weighting.

INMET Instituto Nacional de Meteorologia.

IPCC Intergovernmental Panel on Climate Change.

IPVS *Índice Paulista de Vulnerabilidade Social*.

IQR Inter-Quartile Range.

LAD Local Authority District.

LMIC Low and Middle Income Countries.

LSOA Lower Super Output Areas.

LST Land Surface Temperature.

LUR Land Use Regression.

MASP Metropolitan Area of São Paulo.

MCC Multi-City Consortium.

MMT Minimum Mortality Temperature.

MODIS Moderate Resolution Imaging Spectro-radiometer.

NASA National Aeronautics and Space Administration.

NDVI Normalized Difference Vegetation Index.

NOAA National Oceanic and Atmospheric Administration.

ns Natural Spline.

O_3 Ozone.

OK Ordinary Kriging.

OR OddsRatio.

PM_{10} Particulate Matter of 10micrometers.

$PM_{2.5}$ Particulate Matter of 2.5micrometers.

qAIC quasi-Akaike Information Criterion.

R^2 Coefficient of determination or R-squared.

RCP Representative Concentration Pathway.

RF Random Forest.

RH Relative Humidity.

RK Regression Kriging.

RMSE Root Mean Square Error.

SALURBAL *Salud Urbana en América Latina* - Urban Health in Latin America.

SES Socio-Economic Status.

SK Simple Kriging.

STRK spatiotemporal Regression Kriging.

sza	Solar Zenith Angle.
tv-DLNM	Time-Varying Distributed Lag Non-linear Models.
UGRHI	*Unidades de Gerenciamento de Recursos Hídricos*.
UHI	Urban Heat Island.
UN	United Nations.
USP	University of São Paulo.
WHO	World Health Organization.
WLS	Weighted Least Square.
WMO	World Meteorological Organization.
ZEIS	*Zonas Especiais de Interesse Social*.

Acknowledgments

To my incredible team of supervisors: Dr Daniel Fecht, Prof Paolo Vineis, Dr Monica Pirani and Dr Fred Piel – this thesis was made possible thanks to your invaluable advice, support and guidance. Special thanks to you, Daniela, for being the rock that has held me still when everything else was shaking. Thanks also for believing in me and my ideas, and for encouraging me to pursue them. For all that, I will forever be grateful. Paolo, thanks for the wisdom and knowledge, and for making sure that I never forget 'the bigger picture'. Your dedication to public health and epidemiology are undoubtedly, an inspiration. Monica, I appreciate your expertise in all statistical matters and your patience with me. Your guidance has been invaluable in my research journey. Fred, thank you for your support and guidance during my time at Imperial.

To my collaborators, Prof. Adelaide Nardocci, and PhD assessors, Prof. Antonio Gasparrini and Prof. Majiid Ezzati, your contributions have been paramount. Adelaide, your insights about São Paulo, assistance with ethics approvals, and data access were crucial. Antonio and Majiid, your critical thoughts and invaluable recommendations during my ESA and LSR were instrumental in shaping my PhD into what it is today. I am also grateful to my funders, the PhD President's Scholarship, and Imperial College London, whose support made this dream a reality.

To European Space Agency ϕ-lab - thanks for the opportunity. In particular, thanks to Dr Rochelle Dos Santos Schneider for believing in me and my proposal to do this research visit. Those were some of the most enriching months of my PhD. Special thanks to Giuseppe Borghi and Pierre Philippe Mathieu for sharing their wisdom on earth observation, and for inviting me to create a mural for the ESA ϕ-lab, it was an honor. To the ϕ-lab team, and specially Ali, Casper and Alessandro, thanks for your help and for sharing your enthusiasm for earth observation and machine learning with me. And to the rest of the ESA and Rome crew, Romain, Charlotte, Albin, Stef, Javi and Inés, thanks for letting this random, enthusiastic and hyperactive girl become part of your lives.

To my colleagues and friends at Imperial: Annalisa Sheehan, Charlie Roscoe, Sophie Hamilton, Honor Bixby, Weiyi Wang, Sierra Clark, Emily Muller, Barbara Metzler, and Dylan Wood. Thanks for the chats, laughs, coffee and lunch breaks; these moments kept me going. This PhD experience would have been much less fun without you all! To the *Imperial girls* in particular, you were my rock in moments of despair, always there to give advice and provide comfort. Annie and Charlie, brainstorming and working with you two has kept my enthusiasm for research more alive than ever. Finally, to all the brilliant researchers I've met along the way, thank you for the valuable lessons.

To my London family: Tom, Mimis, Ioana, Gareth, Javi, Noe, Stephanie, Brad, Elina, Arthur, Salomé, Will, Myles, Suzy, Ashlyn, Marco and the girls. Thanks for making London feel like home. A special thanks to you Tom and Mimis, for truly being amazing friends, getting me out the house and

making me smile. Life would be so much more boring without you! And finally, thanks to Imperial College Outdoor Club for bringing me to this bunch and showing me the wonders of the UK.

To the girls: Joanna, Lidia and Micol - thanks for everything, you literally revolutionized my life! We moved in to a house and we made it our home, full of love, compassion, support and laughter. Together we have nourished strength, resilience, compassion and kindness in each other, never loosing sight the joy for play and adventure. Lidia, thanks for sharing your dedication and love for the planet and its people with me, it has been an inspiration. Micol, thanks for being a guiding light with your strength and assertiveness. And to you Joanna, thanks for being the best friend I could have ever asked for! Your crazy energy, love and kindness are medicine for my soul.

To my friends back home and abroad: Melodi, Eva, Marta, Mariona, Mireia, Jordi Moya, Jordi Masnou, Carles, *les nenes de la Serra*, and Hassan. For bearing with my poor communication and late responses. For always making time for me and my express visits home. I cherish every second spent with you and always carry you in my heart. Special thanks to you, Moya, for the late night calls that leave me feeling lighter, for always making me laugh, and for reminding me to take things less seriously. And to you Melodi, for being the amazing superwomen you are and reminding me that I can do whatever I set my mind to. Can't wait to see more of this beautiful world with you!

And most importantly, to my family – simply, thanks for everything. *Mama* and *papa*, thanks for always encouraging me to find my own path, and supporting my decisions. Despite I wasn't able to experience this journey by your side, your love, support and advice have kept me "afloat and swimming". Ivet, thanks for being the best not-so-small sister in the world, and for always having the right advice – you know me more than myself sometimes, which is scary! Can't wait to spend more time with you, and adventure together. *Iaia* and *avi*, oh do I wish you had been here to see me reaching this milestone, but I have felt your love and support through every ray of sun and drop of rain. *Tieta* and *oncle*, thanks for your unconditional love and support.

Lastly, I want to thank me. For showing up, for not giving up even in the dark times, for standing for myself, for being compassionate and kind to myself, for believing in me, for learning that *I am enough*, and most importantly, for never loosing that spark and that passion that brought me to public health research to begin with! This has been a wild ride but also an amazing one. Can't wait to see what is awaiting in the future.

Finally, thanks to YOU for showing interest in my research and reading my thesis, enjoy!

Chapter 1

Introduction

1.1 Rationale

Among the diverse and complex hazards associated with climate change, variations in temperature are amongst the most direct [1]. The undeniable link between temperature and human health is supported by biological and epidemiological evidence[2–8]. Essentially, exposure to higher than optimal temperatures puts a strain on the body which is trying to maintain the optimal internal temperature through thermoregulation. Exposure to extreme temperatures, and/or an impaired or insufficient thermoregulation response, causes physiological damage and, in severe cases, may lead to death [2]. The degree of impact on a population depends on their exposure levels and underlying vulnerability[1]. Adaptation refers to the ability of altering once risk profile by modifying either the exposure or the vulnerability. This adaptation capacity is often tightly linked to physiological, behavioural, infrastructural, and technological factors and thus, varies over time and space.

For instance, the optimal temperature or MMT differs significantly across various climatic regions [9, 10], closely tied to the average daily mean temperature. This suggests that populations can adapt to regional temperatures [11]. Similarly, for a given temperature, the resulting impacts on populations' health have been found to vary. This heterogeneity is most likely a result of the distribution of vulnerability factors. For example, the elderly or those with diabetes have an impaired thermoregulatory system, making them more susceptible to temperature-related death[12, 13]. External factors, including access to services and technology (for example, air conditioning), may determine the adaptation capacity of individuals [12, 13]. The prevalence of these factors are dynamic in time, space and by population group. Despite growing evidence, the temporal variation of the temperature-mortality association and the role of individual and contextual characteristics in defining temperature vulnerability remain understudied, particularly in Low and Middle Income Countries (LMIC) [12, 13].

There is a particular scarcity of studies investigating spatial, temporal and population group-specific temperature-related health risks within cities. More than half of the global population live

in urban areas. Urbanization is a known driver of higher temperatures recorded in urban settings compared to rural surroundings, the so called urban heat island [14]. Micro-climates have also been identified within the city, with localized heat and cold islands [15]. Most studies investigating the temperature-related mortality risk in cities use spatially coarse temperature estimates, such as measurements from one reference meteorological station, thereby loosing information on spatial variability in temperature. This potential exposure misclassification might bias results or smooth the effect sizes towards the null. There is, therefore, an imperative need for highly spatially resolved temperature estimates which are able to capture the spatial heterogeneity of temperature exposure within cities without compromising the temporal granularity (such as daily). Despite growing research exploring temperature-mortality association for population groups and small areas [16–19], it remains primarily in high income countries, and often, at a too coarse resolution for intra-city studies [20, 21].

The spatial and temporal distribution of exposure and vulnerability, and their impact on the temperature health burden, are unique to each city. With around 20 million inhabitants, São Paulo metropolitan area is the most populous area in Brazil, and the fourth most populated metropolis in the world [22]. More than half of the population (11.7 million) is concentrated in the municipality of São Paulo, where it's capital city is [23]. This makes this municipality a good case study to investigate intra-city patterns, otherwise impossible to assess due to statistical power constrains. Moreover, São Paulo is a vibrant city with strongly ingrained social and urban disparities, an heterogeneous urban fabric and strong Urban Heat Island (UHI) [24], which makes it an interesting setting to assess spatial patterns in the effects of temperature on mortality. By delving into the spatial and temporal distribution of factors influencing temperature health burdens, this research can offer valuable insights to inform tailored public health strategies and interventions.

1.2 Aims and objectives

The aim of this thesis is to provide a comprehensive description of the mortality burden of non-optimal temperatures in São Paulo (Brazil), including the exploration of temporal trends in the mortality risk, the description of spatial patterns in the temperature exposure within the city, and the identification of individual and contextual factors that alter the vulnerability to effects of temperature.

In order to achieve this, specific objectives include:

1. To describe trends in the temperature-mortality association, overall and by population group.

2. To build and implement a model to estimate daily mean temperature at high spatial resolution.

3. To investigate how the effects of temperature on mortality vary by different individual and contextual factors.

1.3 Structure of the thesis

In Chapter 2, I critically review relevant literature on temperature-mortality association and the methods used to assess its temporal and spatial variability, in order to provide the reader with an up-to-date overview of the state-of-the-art research. In Chapter 3, I investigate the temporal changes in the temperature-mortality association in São Paulo state between 2000 and 2018 for different population groups. In Chapter 4, I describe the development and implementation of a machine learning pipeline based on a Random Forest (RF) algorithm to estimate daily mean temperature at a spatial resolution of $500m^2$. I then compare the results and model performance to a traditional multi-linear regression model. Finally, I describe the main characteristics and usability of the final data product. In Chapter 5, I showcase how precise exposure data as that developed in Chapter 4 can be used to explore differences in the temperature mortality association by groups of individual and area-level contextual factors. Data sources and methodology vary substantially by chapter and so, they are described in detailed in each chapter separately. I conclude, in Chapter 6, with an overview of my findings and discussion of the implications from a public health perspective. Finally, I provide some suggestions on potential future research directions. This thesis was conducted in close collaboration with Dr Adelaide Nardocci, my collaborator and adviser from the University of São Paulo (USP). Chapter 4 also benefited from the knowledge and skills acquired during my 3.5 month research visit at the European Space Agency (ESA) ϕ-lab in Frascati, Italy, under the supervision of Dr. Rochelle Schneider. For more information on the University of São Paulo (USP) collaboration and European Space Agency (ESA) research visit, refer to Appendix Section 6.7.2 and 6.7.3.

1.4 Ethics and governance

The study was approved by the University of São Paulo ethics committee (CAAE: 31664820.0. 0000.5421) and in accordance with Brazilian Law 13,709/2019 on personal data protection.

1.5 Data sharing

In addition to the traditional scientific publications and conference presentations, the computer code for replicating the results and data (when possible) can be accessed through my GitHub repository[25]. I also built a website for the project [26], where I translated (and continue to translate) the published research into compelling stories with interactive visualizations of the results in plain English. With this, I sought to maximize the dissemination and impact of my research in a diverse and inclusive manner and to reach not only to the scientific community but also stakeholders, policymakers and the general population. For more information, refer to the Annex Section 6.7.1.

Chapter 2

Background

Climate change is altering weather patterns globally, including a progressive increase in the occurrence of moderate to extreme ambient temperatures. The world is now on average a little more than 1°C warmer than pre-industrial times [27, 28], although there is substantial geographic heterogeneity as some populations are experiencing temperatures above 2 °C, while others have experienced relatively little change [29]. Changes in the temperature profile are the most direct and deadly consequence of climate change described to date. For human populations, this poses an enormous threat as exposure to non-optimal temperatures is associated to excess morbidity and mortality. An estimated 5 million deaths per year globally are associated with non-optimal temperatures, accounting for 9% of all deaths [30]. Despite this, the United Nations (UN) Sendai Framework for Disaster Risk Reduction approved 2015-2030 [31], the largest international effort to guide the management natural and human-made disaster, did not include extreme temperatures as one of the listed disasters to prioritize, lessening its importance. Yet, since its approval, mounting evidence has shown the risk associated with prolonged exposure to moderate heat and cold, as well as the short-term impacts of extreme temperature events such as heatwaves and cold spells [3–8]. Thus, mitigating its impacts has now been recognized as a priority by several international organizations including the World Health Organization (WHO) [32] and the Intergovernmental Panel on Climate Change (IPCC) [1]. Here, I focus on temperature and the risk it poses to human health.

2.1 Disaster risk: the case of non-optimal temperatures

Risk has been discussed and defined for years in several disciplines. For consistency and adequacy to the research topic under investigation, hereafter, I employ the terminology used in the climate change community and formalized by the IPCC [1]. According to this terminology, *disaster risk* is expressed as the potential of loss of life, injury or damaged goods resulting from a disaster or disease, affecting a specific community or system at a given period of time. This term can be used to describe the

probability of damage to material objects, ecosystems, and humanity because of an event, either a natural or human-induced disaster.

Disaster risk results from the interaction of three components: *hazard*, *exposure* and *vulnerability* (Figure 2.1)[1]. A *hazard* is a natural or human-induced physical event with potential adverse effects on populations which are *exposed* and *vulnerable*. In this work, the hazard of interest is temperature. *Exposure* refers to how the hazard is experienced by the population of interest as a result of the characteristics of the ecosystem being affected. For example, the physical, political, economical and social characteristics of a setting influence the experience of a hazard shock. *Vulnerability* refers to the likelihood of a community to suffer from the consequences of a hazard shock, causing larger damages and losses. This disproportionate burden can either be an inherent physical predisposition (e.g., genetic), often referred to as *susceptibility*, or external, when it is related to characteristics of the social and environmental processes in place (e.g., occupation or Socio-Economic Status (SES)). Each of these components are essential, yet not sufficient for a hazard to result in impacts. For example, one can live in an earthquake prone region (exposed) but not experience any earthquake (hazard). In this scenario, the vulnerability of the population remains invisible. However, in the event of an earthquake (hazard), the vulnerability of the community will be made apparent. For example, if the buildings were designed to withstand earthquakes (reduced vulnerability) the damages will be lesser than if they were not earthquake-proof (high vulnerability). Thus, a disaster could be seen as an unusual strain to the system which often exposes its weaknesses and strengths.

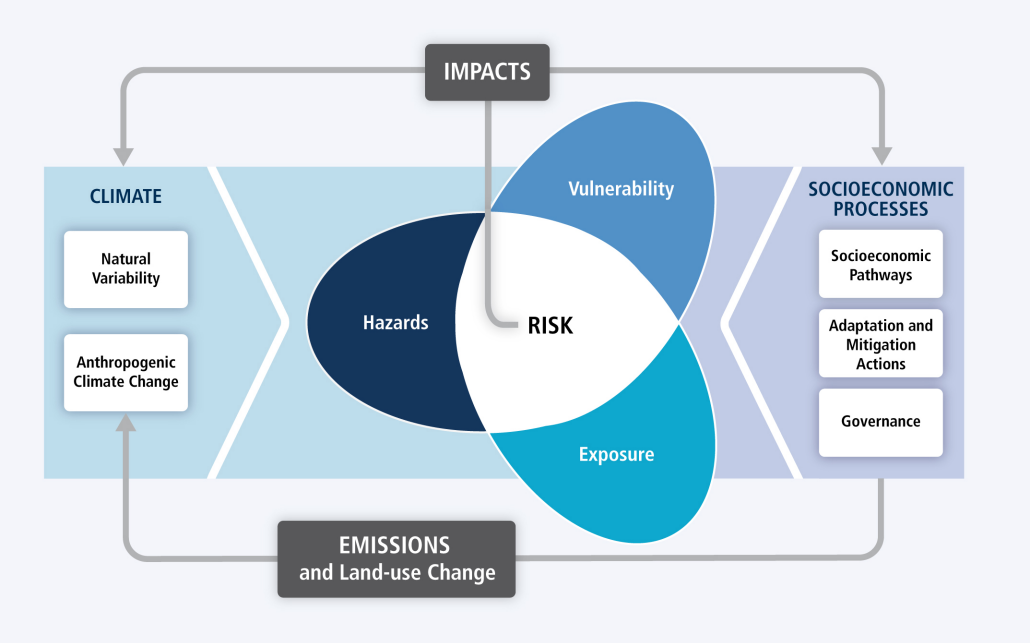


Figure 2.1: **Disaster risk conceptual framework.** Conceptual framework to describe disaster risk. Source: IPCC AR5 Figure SPM.1 [1]

2.1.1 A dynamic and heterogeneous phenomenon

The risk experienced by a system varies enormously across populations, between physical locations and over time, making it a dynamic and heterogeneous phenomenon. This variability is a direct result of the constant evolution of its three components (i.e., *hazard*, *exposure* and *vulnerability*). For example, modifications in the frequency, magnitude and duration of a hazard will have an impact in the resulting risk profile. Alterations in the hazard are often slow-paced, with their changes apparent in the scales of decades and centuries. A clear example of this is climate change. In 2017, we registered a human-induced global warming of on average +1°C above pre-industrial times, a value that is expected to reach 1.5°C by 2040 if this pace of warming continues [1]. Reductions in the magnitude, frequency and duration of the hazard are also possible and are the focus and foundation of climate change mitigation. In addition to modifications in the hazard over time, spatial heterogeneity is also of interest, as not all areas suffer the same intensity, frequency and duration of hazard.

Vulnerability and exposure have been previously referred to as *multi-dimensional, differential, scale-dependent* and *dynamic* systems [33]. The multi-dimensional and differential properties allude to the heterogeneity that exists between locations and across population groups. They are also sensitive to variations depending on the unit of interest, as it is not the same to look at individual vulnerability compared to vulnerability of an entire system, community or place. Although the connection between certain determinants of risk is known, their variability across different geographical units and population groups remains less studied, in particular for fine spatial scales such as intra-urban scales. Finally, vulnerability and exposure are dynamic as they change over time.

Changes in the vulnerability and exposure profiles can result from physiologic changes (*intrinsic adaptation*) occurring as a response to a changing environment, and/or from external factors (*extrinsic adaptation*) such as economic development, behavioral changes, infrastructure improvements, technological advances, preventive initiatives, or improvements in the health-care services, among others. Climate change itself, can act as a catalyst for these changes by adding pressure to the system. For example, it can act as a source of material damage and loss (e.g., damaged infrastructure, displacement), mental health problems (e.g., stress associated to displacement), economic strains (e.g., poverty), or food and water insecurity (e.g., floods and droughts). These stressors, often the aftermath of a disaster but not exclusively, can increase the vulnerability and exposure of a community to a hazard, both short and long-term. Ultimately, the overall risk will likely be the combination of multiple factors of different intent (planned or non-intended), nature (proactive or reactive), time-scales (short-, mid- or long-term) and an effect likely to be heterogeneous across population groups, areas and time. Thus, gaining insights on the particularities of their behaviour is vital to provide a relevant and accurate quantification of the risk at the needed timescale, geographical area, and for the target population of interest.

The IPCC defines *climate change adaptation* "in human systems [as] the process of adjustment to actual or expected climate and its effects, in order to moderate harm or exploit beneficial opportunities"[1]. This term was coined with the intention to be used as a foundation on which to define targets and monitor progress in several national and international agreements. For this reason, this definition focuses on reactive and proactive planned interventions, for it gives a sense of agency and empowerment. Yet, adaptation also includes evolutionary/physiological adaptation, primarily through phenotype plasticity, for which there is abundant evidence [34]. Unfortunately, when investigating adaptation in observational epidemiology, the focus of this work, can rarely distinguish the underlying intent nor the independent contribution of each factor involved. Nonetheless, we can describe whether risk profiles are different over time, across settings and/or population groups. By doing so, we are providing relevant information on the capacity of the population to withstand a hazard, and are giving some insights on the potential drivers of such response.

In the following sections, I explore in detail what we know (and don't know) about the impacts, exposure and vulnerability which define the risk of non-optimal temperatures for human populations.

2.2 Impact: the health burden of non-optimal temperature

2.2.1 Physiological pathways

Temperature homeostasis is of paramount importance in human physiology and ultimately, survival. In the clinical setting, the optimal temperature for prime physiological performance, referred as *the body core temperature*, is between 36.5-37.5°C[2]. Small deviations of one or two degrees, referred to as *normothermia*, are tolerable and cause minimal disruption to cellular function and organ physiology. The human body maintains normothermia thanks to an efficient thermoregulatory response in place which corrects any deviation from the optimal temperature range. Upon its activation through peripheral and central thermo-receptors, the hypothalamic thermoregulatory center, which is located in the pre-optic area of the hypothalamus, initiates a physiological stress response aimed at maintaining the body core temperature within the optimal range. This is achieved by: (i) modifying the evaporative heat loss through sweating and respiration, (ii) up or down-regulating peripheral vaso-dilatation to modify skin blood flow and temperature, and (iii) altering the metabolic heat production. However efficient, when the body is exposed to extreme ambient temperatures that cannot be compensated through internal thermoregulatory processes, life-threatening situations may arise.

Exposure to prolonged and extreme cold beyond the body's ability to adapt results in *hypothermia* and eventually, death. Physiologically, upon reduction of the core body temperatures, the sympathetic system tone increases, activating cutaneous vasoconstriction and decreasing the peripheral perfusion to prioritize vital organ perfusion. This occurs at the expense of a decline in skin and muscle temper-

ature. Brown adipose tissue is activated after only 1-2h of exposure to body temperatures below 17 °-C, increasing local heat production. White adipose tissue would also be able to activate a similar path upon hypothermia, as recently discovered [35]. The respiratory system initial response is hyperventilation, rapidly followed by hypo-ventilation as temperatures continue to fall. An increase in CO_2 in blood follows, which eventually leads to respiratory acidosis. In the gastrointestinal system, motility is decreased, leading to acute gastric dilation and distension of the colon and illeus. Pancreatic injuries are also reported in individuals suffering from hypothermia. Renal affectation due to vasoconstriction is less acute than for hyperthermia, and often reversible. As hypothermia progresses, the blood flow and oxygen content reduces. Consequently, the oxygen consumption of the brain and spinal cord reduces to avoid anoxic injury. This causes a slowing of the cognitive responses. Peripheral nerve conduction and deep tendon reflexes are also reduced. This, together with the peripheral perfusion, may contribute to the cold injuries and loss of manual dexterity observed in patients suffering hypothermia. Eventually, heart rate (*bradycardia*) and blood pressure (*hypovolemia*) decrease to life-threatening levels, eventually causing heart failure.

Exposure to prolonged and extreme ambient heat beyond the body's ability to adapt results in *hyperthermia*, which may lead to life-threatening medical conditions such as dehydration, electrolyte disorders, and heat exhaustion and its progression, heatstroke. From a physiological point of view, the response is as follows. In an attempt to increase heat dissipation, the circulatory system increases the sweating mechanism, eventually leading to a dehydrated state, increases the production of nitric oxide to up-regulate vasodilatation, and shifts heated blood from central to peripheral circulation to facilitate the dissipation of heat through the skin. These phenomena cause a state of hypotension which increases the risk of a cardiovascular shock and eventually dead. Individuals with hypertension have an impaired capacity to shift heated blood from central to peripheral circulation and so, suffer from greater risk of heart failure. Cutaneous vasodilatation results in sweating and flushing in the integumentary system. The respiratory system response involves hyperpnea and hypocapnia. *Hyperpnea* refers to taking deeper breaths than usual increasing the volume of air in the lungs. This adaptation is aimed at increasing heat dissipation through evaporation, yet when in excess it yields a raise in the blood pH, causing a respiratory alkalosis. *Hypocapnia* is the term used to refer to reduced CO_2 in the blood, which has been suggested to induce cerebral hypoperfusion, i.e. inadequate blood flow to the brain. Renal damage may also occur due to hypoperfusion and direct thermal damage. In the gastrointestinal system, we observe a vasoconstriction of splanchnic blood flow in order to mitigate the drop in central blood flow caused by the vasodilatation, and renal hypoxia leading to reactive oxygen and nitrogen species being released. Further intestinal injury may occur due to cellular damage. At temperatures above 40°C, cells will suffer denaturing of proteins, destabilization of cellular membranes, and activation of heat-induced apoptosis pathways; temperatures above 49°C

will cause all cellular structures to collapse, inducing almost immediate (<5 minutes) necrosis of the tissues. During this process, the cells release heat-induced cytokines, interleukin and heat shock proteins into the bloodstream which are toxic to several organs. Nausea, vomiting, hepatic injury and diarrhoea constitute a common clinical picture of hyperthermia. The sympathetic nervous system is activated as a response to the stress. Additionally, individuals may suffer from neurological dysfunction due to direct neuronal damage, alterations in the cognitive pathways, ischemia, hypoperfusion or inflammatory processes. Further damage can be caused by metabolites and potentially hazardous substances penetrating the blood-brain-barrier due to its increased permeability with heat.

In conclusion, extreme ambient temperatures are a threat to body temperature homeostasis and can eventually compromise life. The physiological capacity of an individual to tolerate extreme temperatures above the thermal limits is dependent on the effectiveness of their thermoregulatory response. However, even in healthy adults, there is a limit to the capacity of the body to respond to extreme temperatures, leading to life-threatening conditions. This is worsened and accelerated in individuals with compromised thermoregulation, e.g., the elderly or people with certain chronic diseases such as diabetes [35].

2.2.2 Temperature measures

Various temperature measures have been used to quantify the impacts of temperature on population health. Studies which have compared different temperature measures and indices often report that no single temperature measure seems to be consistently superior to others, varying across population groups, region and outcomes [36, 37]. This reflects the complexity and specificity of temperature-health impacts which are likely to depend on factors such as the extremity of the deviation from optimal temperatures, the duration and timing, as well as concurrence with other physiological stressors with an independent, synergistic or antagonistic effect. Thus, the decision of which temperature measures to use eventually comes down to a combination of methodological considerations, data availability, research interest and relevant exposure measures which better capture the investigated effects, and the local context [38].

With climate change taking more center-stage in public health discussions, some studies have attempted to quantify the impacts of long-term temperature anomalies, defined as temperature deviations from a baseline long-term average, on health. Such deviance manifest in the range of hundreds or thousands of years, and thus it is out of the scope of this project, which focuses on ¡20 years of data. Extreme temperature events, such as heatwaves and cold spells, have also gained attention as their frequency, intensity and duration is expected to increase substantially in the near future [1]. Their findings are key to inform early warning and response strategies such as the and HHAP. Existing evidence suggests a significant impact of heatwave on mortality, yet the magnitude of the effect

estimates varies under different heatwave definitions, as well as population and climate region under investigation [39].

However, most epidemiological studies have used purely meteorological indicators such as mean temperature [40, 41]. These 24h estimates, are often selected for their broad availability, easy interpretation and for providing a comprehensive measure that captures both daytime/nighttime and cold/heat temperatures, offering a general representation of daily thermal stress on the population. For this reasons, they are often the preferred measure for global initiatives such as the Global Burden of Disease [42] or forecasting exercises of the temperature-related mortality under various climate scenarios [43, 44]. However, ambient air temperature may not be the best predictor of thermal discomfort, as this is influenced by humidity and wind, among other factors. And so, several attempts have been made to combine these meteorological conditions in one index. For example, *apparent temperature (AT)* , used in numerous studies [45–47], combines air temperature with dew-point temperature (Zanobetti and Schwartz, 2008). Other such indices include the *Humidex* index [48] or the more recently developed wet-bulb globe temperature (WBGT) and universal thermal climate index (UTCI); the two latter, an attempt to have a standardized, universally applicable, physiologically relevant, and appropriate index for use in a range of bio-climatic regions [49]. Their validity across regions and population groups and relevance for epidemiological research remains to be quantified.

Another fundamental consideration is the difference between outdoor temperature and personal exposure. Personal exposure can vary greatly depending on individual behaviors, housing conditions, work environment, and other factors, leading to exposure misclassification when only outdoor temperature measures are used. In an attempt to address this, some studies have attempted to measure personal exposure by using wearable temperature sensors [50] or estimating the time-weighted sum of temperature in the places visited by the subjects, mirroring approaches used in other fields such as air pollution. Nevertheless, obtaining personal exposure data for a large population can be logistically challenging and expensive and it is not possible at a population-level. Several attempts to infer indoor temperature from outdoor temperature have failed to capture the complexity of this association [51]. For this reason, and despite it is an important limitation, outdoor temperature studies are still prevalent in epidemiological research. Finally, public health planning and interventions are often based on population-level data. Understanding how outdoor temperature affects overall mortality rates helps policymakers develop strategies to mitigate the impacts of extreme temperatures on a community-wide scale. For all these reasons, investigating indoor air pollution was outside the scope and budget of this project.

For this project, I selected outdoor daily mean temperature because: (i) the data was available and of good quality across most stations in the study area, as opposed to other meteorological measures such maximum and minimum temperature, humidity or wind; (ii) it captured both heat and cold

events, as opposed to other measures such as maximum or minimum temperature that focus only on one end of the temperature spectrum, as well as daytime and nighttime temperatures, providing a more representative measure of the overall thermal stress on the population; (iii) previous literature in the region show little to no effect of humidity on temperature-mortality associations [52], and so a simpler measures was favoured; (v) it is a widely used indicator in epidemiological research, providing consistence and comparability across studies, and finally, (iv) with daily mean temperature being a single, easily understandable metric is often considered more relevant to public health planning and communication, an important consideration in my research.

2.2.3 Temperature-mortality relationship

Temperature impacts on mortality have been extensively studied [3–8]. Despite the high death toll of extreme events, moderate temperatures explain most of the overall temperature-attributable mortality burden[3]. The effects of temperature on mortality are non-linear, with the highest mortality risks occurring at the extremes of the temperature distribution, and are lagged or delayed, meaning that the impacts expand over several days. In general, the shape of the exposure–response relationships is non-linear, commonly described to be V-, U-, or J-shaped, with the lowest point or the temperature at which a given population shows the lowest risk of death, referred to as Minimum Mortality Temperature (MMT). That is, the risk of mortality increases as temperatures deviates from the MMT toward the left (cold) or right (heat). The risk associated to a given temperature, calculated as the cumulative effect over all the lags, is known as the Cummulative Relative Risk (cRR). The slope of the curve defines the increment of risk per unit of temperature change; the steeper the slope the larger the increase in risk.

The extent to which the effects of heat and cold on mortality are the results of bringing forward deaths by a few days or week, the so called *harvesting effect* remains unclear as it has been reported in some occasions[53–55], yet not in others [56, 57]. Basically, the harvesting effect is a mortality displacement phenomena characterized by the decrease in the number of deaths below the expected following an initial increase [53, 58]. This is likely due the early death of people whose health is already compromised, resulting in a displacement of mortality. Thus, it often affects the elderly and those with terminal illnesses [53, 58]. In time-series studies, it is not possible to quantify how many years of life have been lost. However, it seems clear that they contribute to the overall excess deaths resulting from temperature, in particular for cold [53, 58]. From a statistical point of view, in order to capture the totality of the excess deaths resulting from temperature, including those displaced, it is important to include multiple time lags (which will be specific for each region). This provides a continuous representation of the risk over long periods, capturing any potential harvesting effect.

2.2.4 Study designs: time-series and case-crossover

Studies assessing the association between temperature and mortality have mostly used two study designs: *time-series* and *case-crossover design*.

The time series approach uses a series of mortality counts and temperature recordings at evenly spaced time units (most commonly daily due to the short-terms effects of temperature) often assuming an over-dispersed Poisson regression model [59]. Time-varying confounders can be added to the model and adjusted for, such as air pollution, seasonality or long-term trends. This approach has some limitations. For example, when it comes to assessing differences across population groups or spatial units, the data has to be stratified, causing small statistical power or in the worst case scenario, model non-convergence. Some have addressed the lack of statistical power by applying a two-step approach where first univariate time series are investigated and then the results are pooled together in a multivariate meta-analytical model [60, 61]. However, this approach looses information of the individual constituents in favour of a combined estimate.

The case-crossover design, first described by Maclure in 1991 [62], is a special case of of time series analysis which uses a case-only approach used to assess the effects of short-term exposures, such as temperature. In this design, each event is matched with one or multiple control events for the same subject at different times when the event did not occur. It follows that time-invariant (individual or collective) subject-specific characteristics are adjusted by design. This leaves only time-dependent covariates to adjust for in the modelling stage. Several strategies for choosing control days are proposed, yet the most commonly used in the temperature-mortality literature is the *time-stratified case-control matching*, which matches cases with controls based on certain temporal constraints [63]. The most commonly used criteria states that for each case, three to four controls must be selected that fall within the same day of the week, month and year (see Figure 2.2). By selecting controls within a short time window around the event date, we ensure that we control for seasonality and long time trends, as shown by [64]. From a statistical point of view, the association is the assessed through a conditional logistic regression. Levy and Lumley[65] showed that a time-stratified approach to choosing controls resulted in a proper conditional logistic likelihood, and Bateson et al., 2001 [66] demonstrated with simulation studies that this approach gave unbiased effect sizes and coverage probabilities even with strong seasonal confounding. This design allows for the use of exposure estimates specific for each individual, and thus, it has been used specifically to investigate differences in the risk across groups, either defined by individual or contextual characteristics [17, 36, 67–72] .

When compared, these two approaches seem to yield similar risk estimates or slightly more conservative estimates for case crossover studies than time-series studies [45, 73, 74]. It has been suggested that controlling for seasonality and long-term trends through the design itself, as done by the case crossover design, may be advantageous. Nevertheless, other studies have found that estimates

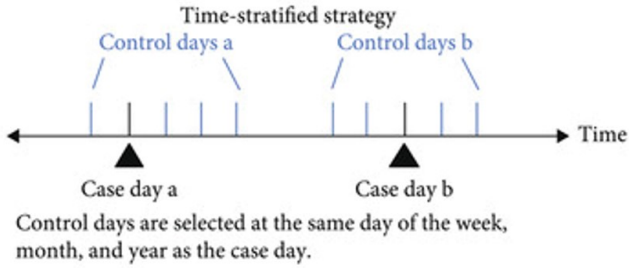


Figure 2.2: **Time-stratified case-control matching.** Diagram showing how controls are selected for each case based on day-of-the-week, month and year matching. Source: Figure 1 in [63]

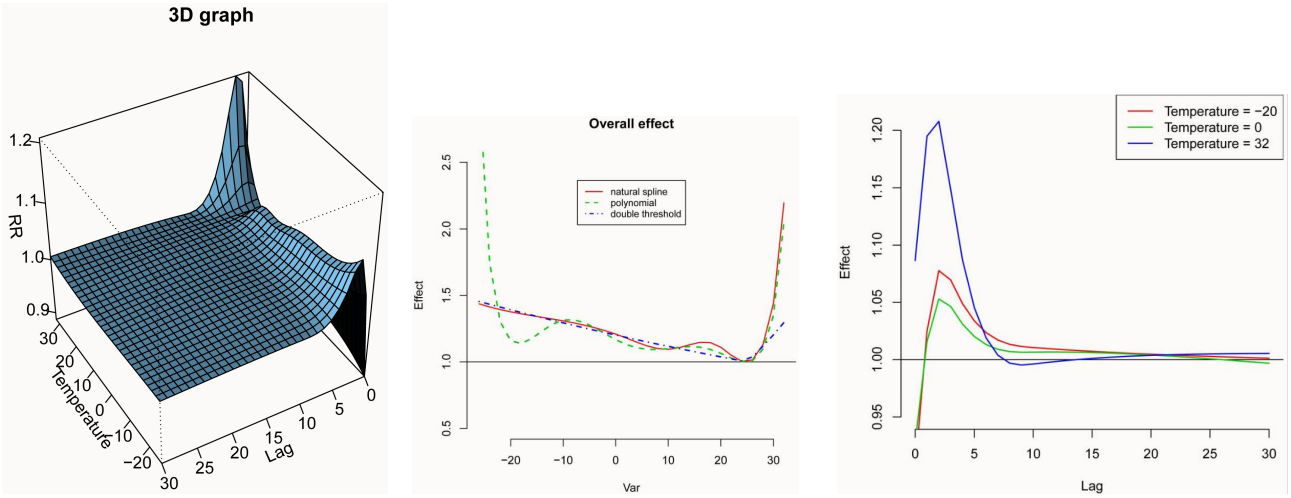
from both case-crossover and time series models were consistent and comparable [71].

2.2.5 Accounting for non-linear and delayed effects: the DLNM framework

In the literature, there are several approaches that address the non-linearity of the relationship, such as through the use of splines or polynomial curves [75]. Similarly, several methods have been used to take into consideration the delayed effects of temperature, for example, by averaging across multiple lags or modelling the lag as a spline[75]. The Distributed Lag Non-linear Models (DLNM), formalized by Gasparrini A. in 2010 [76], is a class of models which allow for complex non-linear and lagged dependencies at the same time by using a bi-dimensional space of functions. One dimension describes the exposure-response association while the other describes the lag-response association. These basis functions are combined through a tensor product, resulting in a bi-dimensional crossbasis function which captures the exposure-lag-response association. The inclusion of a lag dimension, means that one can explore how the temperature of the days prior to the event (i.e., lags), impacted the event. This is particularly useful when investigating cold, as its effects expand over long periods of time, and to account for the *harvesting effect*.

Figure 2.3 provides an illustrative example of the output obtained from fitting a DLNM model to temperature-mortality data, taken from [77]. In (i), we can see the risk represented as a three dimensional surface where the risk is shown as the cRR (*z-axis*) over the continuum of temperature (*x-axis*) and the lag dimension (*y-axis*). The effects can also be represented as the cumulative effect of all lags over the exposure continuum (exposure-response curve, Figure 2.3(b)) or the cumulative effect of the exposure over the lag distribution (lag-response curve, Figure 2.3(c)). The exposure-response curve can be interpreted as the total effects observed for a given temperature taking into account the contribution of all the lag periods preceding the event; the lag-response is the risk associated to each lag for a specific temperature defined by the modeller. In Figure 2.3(c), each color illustrated the lag-response curve at a specific temperature value shown in the legend.

When specifying the cross-basis function for the DLNM model, there are three main elements that need to be specified: (i) the class of basis functions, (ii) the number and location of knots for



(a) 3D risk surface (b) Exposure-response curve (c) Lag-response curve

Figure 2.3: **Examples of visual representations of DLNM outputs.** (Left) Three-dimensional plot illustrating the exposure-response relationship between temperature and all-cause mortality over the temperature (x axis) and lag dimension (z-axis) with reference at 25°C. (Middle) Exposure-response curve modelled using different basis functions. (Right) Lag-response curve at different temperatures. Source: Gasparrini et al., 2011 [77]

each basis function (which determines its flexibility), and (iii) the maximum lag to consider. The selection of the basis function is primarily informed by the literature and the biological plausibility of the shape. It has to be a compromise between complexity, generalizability and ease of interpretation. For the exposure dimension, the most commonly used basis functions are quadratic β-spline and cubic β-spline, and for the lag dimension, cubic β-spline. There are two approaches to decide the number and placement of knots, which is equivalent to defining the Degrees of Freedom (df), for the basis functions. On the one hand, one can use agnostic tools to identify the best combination of number/placement for the data, for example, via generalized Cross-validation (CV) [78]. Alternatively, one can assign pre-defined placement and number based on the literature and/or plausibility. Finally, the maximum number of lags to be modelled needs to be defined. Again this can be defined agnostically or informed by the literature. Most often, the chosen approach is a combination of pre-defined, literature-informed parameters and agnostic testing, in which we use the literature to narrow down the range of feasible parameters and then select the best combination based on model performance.

Since it's formalization, the DLNM modelling framework has gained users [] and it is currently amongst the state-of-the-art methods of choice to investigate the impacts of temperature on health [12] for both time-series [40, 79] and case-crossover study designs[67, 69, 70, 72, 80–82]. A full description of the DLNM modelling framework, algebraic notation and applications can be found in Gasparrini et al., 2010[76]. Recently, Gasparrini et al., 2022 [16] proposed a new approach based on an extension of the two-stage DLNM framework for case time series design using nested spatial units and adding a downscaling procedure. A full description of the method can be found in [83] and in [16].

2.2.6 Temporal variability in temperature-related health impacts

A growing number of studies are investigating temporal trends in the temperature-mortality relationship. Several such studies have reported some degree of adaptation to increased temperatures for many locations [11, 40, 84–87], mostly in developed countries. Fewer and less conclusive studies have reported adaptation to cold temperatures [86]. However, the magnitude of the adaptation varies substantially by setting, highlighting the importance of the local context and the need for city-specific assessments[13, 88]. Differences are also observed by population groups such as age and sex, exposing disparities in adaptation capacity[84, 89–91]. Evidence on other population groups such as ethnic or socioeconomic groups is limited [86, 92, 93].

In the last decades, studies assessing temporal changes in the temperature-mortality association have been on the rise [40, 89, 94–101], yet they suffered from several limitations. Firstly, a large proportion of studies assessed changes over time by dividing the study period in either overlapping [89] or non-overlapping time periods [102], substantially hindering the statistical power to detect differences. Studies on cold have been limited and shown conflicting and inconclusive results [41, 97, 103], but as populations adapt to heat, it is important to assess whether this is in detriment to an increased susceptibility to cold [41, 103]. Furthermore, there exists a geographic disparity in the availability of evidence on trends [104], with most studies set in the USA [96, 98], Asia [97, 99, 105], and Europe [89, 92, 94, 101]. Some studies have explored multiple cities [100], yet they are dominated by cities in Western countries. Finally, most city-level studies depict the effect on the whole population [85], with only few investigating effects by age and sex [84, 89–91], and even less so for vulnerable populations, such as ethnic groups, education level and/or socioeconomic status [86, 92, 93]. This is mainly due to a lack of individual level information and poor statistical power due to small number of cases. Some studies have attempted to address the latter by using pooled estimates across cities [100], in detriment to capturing the individuality of each city. With more than 11.3 million inhabitants, São Paulo is the largest city in Brazil and suffers from both the effects of cold and heat [106–108]. To date, no study has assessed the local temperature adaptation patterns in São Paulo except for a multi-city study which pooled the estimates for multiple cities in Brazil together [103].

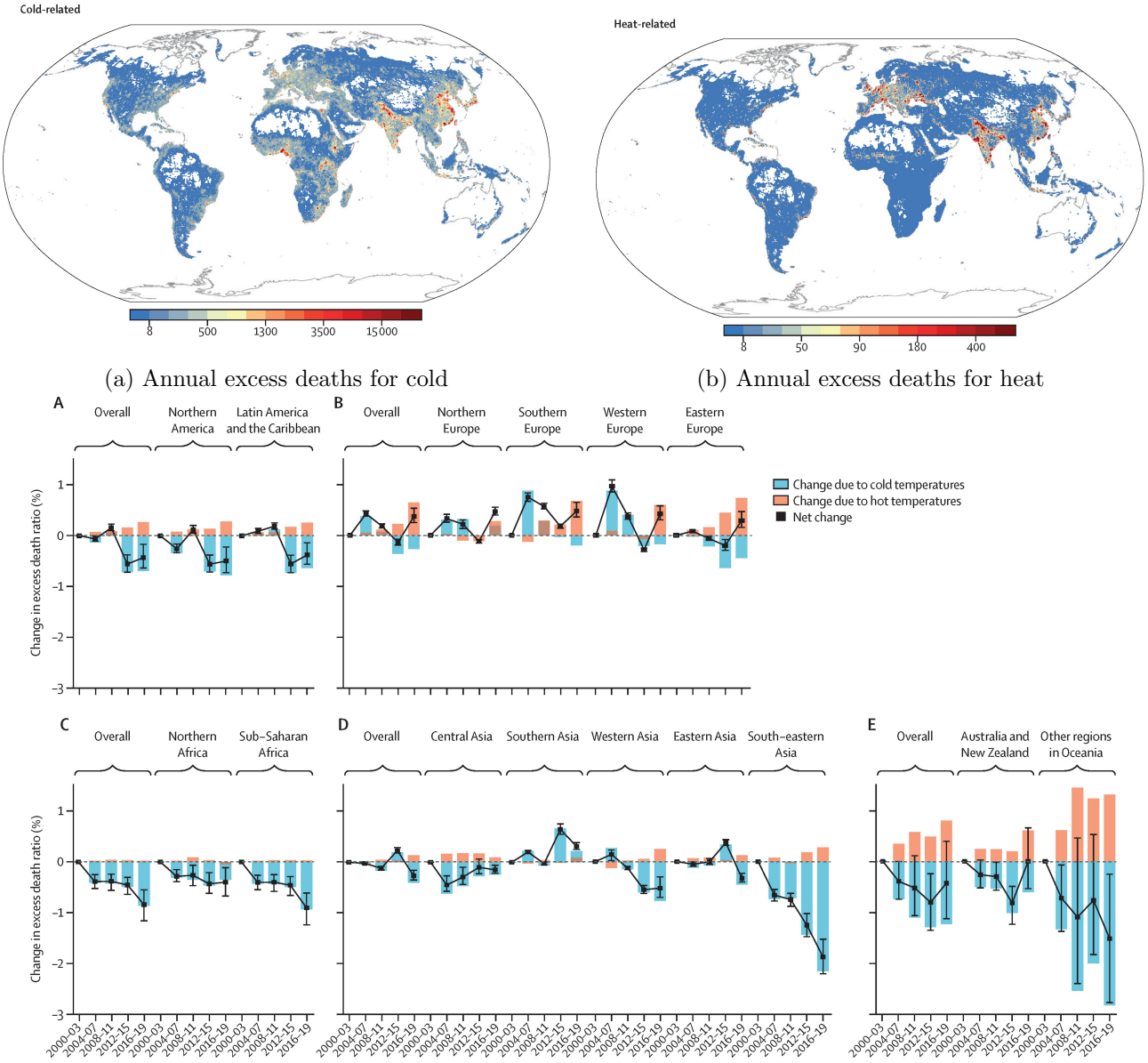
2.2.7 Spatial variability in temperature-related health impacts

The MMT is an important indicator of adaptation as it captures the optimal temperature at which a given population is best suited for survival. Evidence from different locations [9] show that this value varies substantially by population and region, capturing long-term adaptation to local climate. In 2019, Yin Q and colleagues [10] estimated the global distribution of MMT at a resolution of $0.5^\circ \times 0.5^\circ$ grid for 2010. They used temperature and mortality data from 420 locations covering

six continents to estimate the regional MMT and then applied a multi-linear regression with seven variables available at the grid level. These coefficients were then used to infer the MMT at each grid. A similar approach was taken by Zhao and colleagues in their 2021 study [30] , where they provided the first ever global overview of the spatiotemporal heterogeneity in the mortality burden attributable to non-optimal temperatures, defined by the authors being above or below the MMT. To do so, the authors explored city-level mortality and temperature data across 750 cities in 43 countries globally, as part of the Multi-City Consortium (MCC) [1], between 2000 and 2019. Once the city-specific temperature-mortality associations were obtained, they were combined using a meta-regression together with other covariates available at the same grid-level as the temperature dataset, such as GDP per capita. Finally, the mortality burden attributable to non-optimal temperatures was estimated applying the meta regression coefficients on the grid-specific temperature and covariate values for each day between 2000 and 2019. Figure 2.4a and 2.4b show the annual excess deaths for cold and heat, respectively, as published in [30]. The cold attributable excess deaths for cold vary as much as from 8 to 1500 in some areas in Asia and the African belt. In North and South America, we can see some hotspots with higher excess deaths, mainly corresponding to the location of large cities (Figure 2.4a). For heat the variability is lower, ranging from slightly below 8 and above 400 annual excess deaths (Figure 2.4b) The authors also compared the temporal variability in the risk to heat and cold. They demonstrate that the risk varies over time and that this variation is heterogeneous by region (Figure 2.4c).

In order to protect population's health it is imperative that governments and stakeholders have access to relevant data for their area. From these evidence it is clear that the health burden associated to temperature varies substantially over space and time, making evidence from one region not necessarily transferable to another. This is problematic, particularly for Low and Middle Income Countries (LMIC), as most evidence comes from high income countries. Even when available, studies set in LMIC are majorly skewed towards Asian countries. According to a review from Green et al., 2019 [12], of the 146 studies conducted in LMIC between 1980 and 2018, 103 were conducted in Asia, mainly in China (n=82), whilst only 10, 8 and 7 studies were conducted in Africa, the Middle East and South America, respectively. This includes any study assessing temperature-health outcomes, not necessarily investigating effect modification. Thus, highlighting the need for local and time-relevant studies, particularly in LMIC.

[1]The MCC Collaborative Research Network is an international collaboration of research teams aiming to produce epidemiological evidence on associations between environmental stressors, climate, and health. Full description in this chapter Section 2.5



(a) Annual excess deaths for cold

(b) Annual excess deaths for heat

(c) Temporal variation of heat/cold excess deaths by region

Figure 2.4: **Historical spatial and temporal variability in annual excess deaths for cold and heat.** (Top) Spatial distribution of the annual excess deaths for cold (left) and heat (right) globally using a 0.5 x 0.5 ° grid. (Bottom) Change in annual excess death ratio between 2000 and 2019 compared with the 2000–03 average by region. Source: Zhao et al., 2021[30]

2.3 Hazard: ambient temperature

2.3.1 Overview

Ambient temperature is defined as the temperature of the air at two meters above the land, sea or in-land waters. The temperature experienced at a particular place is the result of atmospheric processes as well as regional factors. Interactions between the land and the sea/ocean, radiation energy balances or chemical interactions specially in the upper atmosphere, are some examples of the atmospheric processes in place. Time of the day and season, proximity to water, vegetation, wind

direction and altitude, are some examples of regional/local factors which influence the temperature in a place. Both these elements influence the temporal and spatial variability of temperature.

The spatial variability can be explored at different scales. At a global scale, the world can be divided into five climate zones and 30 sub-types based on their temperature and precipitation patterns, according to the Köppen-Geiger classification. Mounting evidence shows that populations have evolved to adapt to their local climate, showing substantially different risk profiles to meteorological hazards. This is also the case for the definition of 'non-optimal temperature' for a given population[9, 30]. In turn, evidence from a given climate region may not necessarily apply to another, thus supporting the need for local and climate-specific studies. At a regional and local scale, temperature shows substantial variability. Several factors have been shown to influence that variability, with land cover being the main driver. In large metropolitan areas where the land use cover changes substantially in small areas, there can be significant spatial variation in the ambient temperature [109–111]. In Section 2.5, I explore this phenomenon and its factors in detail. For the temporal variability, the clearest example is climate change. For temperature, this means a general increase in the average values and in the duration, frequency and magnitude of extreme events [1]. These long-term changes are exposing populations to temperature ranges that are not common for their region, and for which they are not adapted. These modifications in temperature, have been suggested to have triggered physiological and external adaptation[84, 86, 95].

In short, to accurately quantify the risk of non-optimal temperatures, it is vital to have ambient temperature data at an appropriate temporal and spatial scale. Previously, epidemiological studies have overlooked the spatial component which at best, has resulted in a smoothed perception of risk that does not meet the needs of policymakers and stakeholders, specially in metropolitan areas; at worst, it may have bias the findings providing a distorted image of the risk. The lack of temperature data at a spatial and temporal resolution which satisfy regional and local needs for epidemiological studies has been the main limiting factor, and it is still to date [112]. Below, I provide an overview of the temperature data sources and modelling frameworks available, highlighting their strengths and limitations.

2.3.2 Temperature data types

Depending on the approach used to generate or obtain the data, we can distinguish four classes of temperature data: (i) in situ recordings at ground monitoring stations, (ii) climate models and their downscaled local datasets, (iii) reanalysis data, and (iv) data generated through interpolation models. Below, I review each of them in detail in the context of epidemiological literature on temperature-mortality relationship.

2.3.2.1 In situ recordings

Meteorological stations include *in-situ* recordings at a given time and location captured by meteorological equipment on the ground. Some of the major inconvenient of using meteorological stations to assess temperature exposure include: (i) point samples are not optimal to assess the spatial variability as they only capture the immediate surroundings; (ii) they are highly dependent on regional infrastructure regarding their set up and maintenance; (iii) potential temporal paucity of information due to the operation of certain stations being interrupted (maintenance, broken, vandalism, etc.) or ceasing, (iv) they suffer from data quality issues as instruments deteriorate; (v) arbitrary locations which may not capture the heterogeneity of the climate conditions [113], and (vi) in order to capture spatial and temporal variability, a large density of stations is required in space and time. The latter is usually an issue, specially in less developed and urbanized areas, where the lack of resources hinder the availability of such recordings [114]. Thus, this results in an over-representation of temperature-health assessments from urban settings in developed countries, with some exceptions (i.e., China, and other countries of Asia and the Pacific) [12, 13]. To address the lack of temporal and spatial continuity, new products have been released for public use. Broadly, we can distinguish three categories depending on the methods used to generate them, namely numerical climate models, re-analyses datasets and interpolation approaches.

2.3.2.2 Climate models and their downscaled datasets

Numerical climate models are comprehensive and complex mathematical models of the atmosphere, which simulate the behaviour of the climate, capturing vital interactions between the main climate drivers. They are based on primitive equations that follow the basic principles of physics. There are a range of numerical models available for climate and weather forecasting, each with different specifications depending on their purposes. The primary challenges of these models are the computational requirements. To reduce said computation expense, fine resolution models are only run for specific and defined purposes, hence not available globally nor for long time periods.

Local downscaled climate models attempt to overcome this drawback. They can be generated through downscaling of global numerical models or reanalyses data downscaling, or through interpolation methods. Downscaling refers to the process of bringing originally coarse gridded climate data to a finer spatial resolution suitable for local studies. *Dynamic downscaling* uses a global climate model as boundary conditions and applies physical principles to reproduce local conditions. They resemble numerical climate models in that they apply the laws of Physics to describe the climate processes in place. Depending on how fine of a resolution they can reach, they are known as regional (25 to 50km^2) or mesoscale climate models (1 to 5km^2). The smaller geographical extent and consequently lower complexity of the model, make the computational expense of these models smaller than global numerical models. Nonetheless, these are still highly technical and computationally intense approaches, making them inaccessible to all. *Statistical downscaling*, on the other hand refers to the use of statistical methods to describe relationships between climate variables from large-scale climate models and local climate responses. Statistical downscaling produces relatively similar results to dynamic downscaling when it comes to historical and present climate; however, it is less appropriate for forecasting as future local climate conditions may not resemble historical and current ones, which are used to train the models. Both global and local climate models describe general meteorological patterns, yet they do not necessarily describe actual ambient temperature records experienced locally, which is essential to assess health impacts.

2.3.2.3 Reanalysis data

An alternative to numerical climate models that requires less computational resources and takes into account ground level ambient temperature records, are *reanalyses datasets*. In general, they combine global or regional weather forecasting models constrained to historic observations from meteorological stations via data assimilation to estimate historic temperature over a continuous surface and time period globally. They are constantly revised ensuring a high quality of the data and undergo periodical upgrades. There are different available reanalyses datasets, yet the European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalyses datasets [115] have received particular interest among the epidemiologists community. This global, spatially resolved dataset is freely available from the Copernicus Climate Data Storage [115] and covers a wide range of atmospheric variables, including temperature and humidity 2-m above the ground, as well as other variables such as wind speed or atmospheric pressure. Currently, Copernicus offers two global data products: the ECMWF Re-Analysis 5^{th} generation (ERA5) dataset [116], available at 31km^2, and the ERA5 -Land dataset[117], available at 9km^2. Certain regions have developed specific regional reanalyses datasets and provide access to the data freely, as it is the case for Europe and the CERRA dataset developed by the Swedish Meteorological and Hydrological Institute and available at 5.5km^2[118]. The coarse spatial

resolution of these re-analyses products leads to the averaging of spatial variability, omitting micro-climatic areas. This is particularly problematic for small-area epidemiological studies, particularly those investigating short term impacts of environmental variables in densely populated areas with high spatial heterogeneity, as it is the case of cities.

The need for appropriate spatiotemporal temperature data and the arrival of new data sources, has led many researchers to explore new methods to retrieve temperature data at high spatiotemporal resolutions, mainly interpolation methods such as regression and geostatistical models, and most recently, machine learning techniques.

2.3.2.4 Interpolation methods

Let us take an environmental variable of interest, such as ambient mean temperature, for which we want to predict the values at points for which we do not have recorded data. We expect said variable to vary over time and space. If we are to make predictions for locations for which we do not have measurements, it is imperative to capture the underlying spatial and temporal patterns that describe its spatial and temporal variability, so that we can make robust inferences. We can generalize (Eq. 2.1) that the spatiotemporal variability of a given target variable is composed of: (i) a deterministic component, also known as trend or fixed terms which can be explained by the behavior of auxiliary variables, and (ii) a spatially (and temporally) correlated random component or stochastic process, which captures the residual spatial (and temporal) autocorrelation, i.e., the variability that we are incapable of capturing through auxiliary variables. Depending on the assumptions made with regards to each of these components, we can distinguish different approaches: deterministic, regression and machine learning approaches.

$$ST\ variability = ST\ trend\ component + Residual\ ST\ correlation \tag{2.1}$$

Deterministic approaches assume that there are deterministic predictors, often location and time, that explain the behaviour of the variable of interest and that there is no residual spatiotemporal component. The most common approaches are interpolations that use the information of nearby points as the main (and sometimes only) deterministic predictor. They builds on Tobler's First Law of Geography by which "everything is related to everything else, but near things are more related than distant things" [119]. The most common is the *spatiotemporal kernel predictor* approach. Kernels are functions that quantify similarities between locations, acting as weights. The simplest form of spatiotemporal kernel interpolation is Inverse Distance Weighting (IDW) [120]. IDW is a weighted average of the data points, giving the closest locations a higher weight. Euclidean distances are usually used to describe the spatial and temporal closeness between observations, and assign the weights.

However, these approaches ignore interactions and associations between the variable of interest and other factors, considering only time and location as the main predictors, which may lead to a poor performance of the model in the best cases, and to bias in in the worst cases. In order to address this issue, some researchers have extended or enhanced this approach, for example by pre-processing the input data to consider lag correlations [121] or coupling with other complimentary approaches [122].

They seem to perform relatively well for variables with stable and homogeneous spatiotemporal distributions and for set ups with a high density of randomly distributed training points that cover most of the variability in the variable of interest within the area of prediction. This can be the case for temperature modelling if considered at coarse resolution and for homogeneous landscapes, yet not so for urban environments with the well-established presence of micro-climates [15]. However, these conditions are seldom met and so, more complex frameworks are needed.

Regression and machine learning approaches address exactly that problem. They generalize that there are covariates other than time and location which are needed to explain the spatiotemporal dependence of the observations. These covariates can capture spatial only, temporal only, and/or spatiotemporal dependence of the observations. In this context, earth observation satellite data is particularly interesting as it provides detailed information of a large number of predictors at a fine spatial and temporal resolution across the globe. Satellites are loaded with different visual, optical, infrared, microwave, radar, airborne and acoustic/near-acoustic sensors able to capture numerous properties of the Earth's atmosphere and surface. One such property is the radiant temperature emitted from the Earth which is used to retrieve Land Surface Temperature (LST). This temperature is highly correlated to ambient temperature, yet not identical. And so, in recent years, there has been a growing body of research exploring different statistical approaches to model ambient temperature from LST and other predictors which can capture the residual spatial and temporal heterogeneity, i.e., not explained by LST. The Moderate Resolution Imaging Spectro-radiometer (MODIS) LST product is widely used as one of the most important covariates for temperature interpolation. Dos Santos et al., 2020 [111] , Sekulic et al., 2020 [123] and Kilibrada et al., 2014 [124] used MODIS LST as the main covariate for their models to predict temperature at $1km^2$ for London, Serbia and globally, respectively, just to mention some. Distance to the coast, vegetation indices, water body indices, land cover, and population density are other covariates used in temperature modelling.

In short, interpolation methods, use covariate information together with ground data to identify the existing associations and use those to predict the variable of interest at locations for which there is no observations. There is a wide range of methods we can use to discover and describe these associations. In the next section, I provide an overview of regression and machine learning approaches, two common families of interpolation methods, and describe in detail the ones used in this study: Land Use Regression (LUR) models and Random Forest (RF).

(a) Regression approaches:

Land Use Regression (LUR) models are one of the most used regression interpolation techniques for environmental modelling. They use regression to model the association between spatially (and temporally) resolved data and the variable of interest. This approach poses two main issues. Firstly, it often uses multi-linear regression, imposing the constrain of linear association between the variable of interest and the covariates. This is seldom the case in the real world, where non-linear associations are likely to be present. Secondly, its foundation lies on the generalization that most (if not all) the spatiotemporal variability of a given variable of interest can be modelled. If this assumption holds, interpolation techniques such as LUR, should be able to explain all the spatial and temporal variability, and provide unbiased estimates. This may be the case if we have information on absolutely all the covariates that are associated to its space-time variability to some degree, or for models with coarse spatial or temporal resolution, where the lack of information may be marginal for the purpose of the model. However, this becomes problematic when we are modelling at high spatial and temporal resolutions, such as daily at an intra-urban scale. In this scenario, it is likely that one or more of these challenges arise: there are relevant covariates for which we do not have information, the temporal and/or spatial resolution of the covariates is insufficient to capture the variability of the target, and/or the covariates have measurement error and biases. These challenges cause interpolation techniques to miss part of the spatial and/or temporal variability of the target variable. The spatiotemporal autocorrelation not explained by the model is known as *residuals*.

Kriging is a geostatistical interpolation technique that focuses precisely at modelling the spatial (and temporal) autocorrelation structure of the data. To do so, it builds on Tobler's First Law of Geography [119]. And so, it gives more weight to sample points located nearby than those further away. To define the distribution of weights, we fit a *sample* or *empirical variogram* to the observations. This function provides a description of how pairs of observations are correlated in function of their distance. Finally, the empirical variogram is used as prior information to interpolate values at unsampled points. There are different kriging techniques depending on the assumptions made with regards to the trend component in (Eq. 2.1). Simple kriging assumes that there is a global mean and that this is known to the modeller. Ordinary Kriging (OK) assumes that the mean is an unknown constant which can be estimated from the observations. In both cases, only the stochastic process remains to be modelled as we assume the trend is known or easily estimated from the data. The objective then, is to find the optimal linear predictor or *kriging predictor* of the stochastic process so that the error between observed and predicted is minimized. OK has been used previously to interpolate maximum and minimum temperatures [125]. Nonetheless, both techniques assume the mean is constant over the spatial (and temporal) field, which is an unrealistic simplification of how environmental variables behave in the real world.

A natural extension of LUR, OK, and Simple Kriging (SK) is Regression Kriging (RK) [2], which combines both deterministic modelling of the trend and modelling of the residual spatiotemporal autocorrelation. In this case, one uses a linear function of some auxiliary variables to describe the trend. Different techniques can be used to model the trend, such as linear regression or generalized additive models, just to mention some. Once the trend is modelled, the approach is the same as for the OK/SK, yet here we use the remaining spatiotemporal variation not explained by the model or *residuals*. For modelling purposes, we assume that the residuals are stationary, meaning that the mean and variance of the residuals are constant through the space-time domain and so, they can be modelled through a variogram. In short, RK is a two-step interpolation technique which models the deterministic and the stochastic component separately. In most cases, RK has proven to give better results than Original Kriging [127, 128] and other simpler interpolation methods [129]. Some exceptions include the modelling of variables for which the covariates are able to capture all the variability of the observed variable, in which case the RK would perform similar to LUR, or alternatively, when there is no association between the variable of interest and the covariates used in regression of the trend, in which case RK would perform similarly to a Original/Simple Kriging. Aside from these special cases, RK is regarded as a well-performing and flexible approach to obtain unbiased estimates of spatially continuous variables.

Within RK, *spatiotemporal regression kriging (STRK)* has become popular in recent years. It is a natural extension of RK which, besides the spatial component, it also considers a temporal component and its interaction with space. Its use has been widely increased thanks to the release of several comprehensive packages such as the development of *R gstat* [130] and *spacetime* (r) packages and the extensions to handle spatiotemporal variograms and interpolation integrated by Graler et al., 2016 [131] within the *gstat* package. Kilibarada et al., 2014 [124] and Sekulick et al., 2020 [123] showed that this approach was superior to traditional kriging methods for temperature modelling.

[2]Note: In the literature, RK is often used interchangeably with Universal Kriging (UK) or External Drift Kriging (EDK) as they all assume that the trend varies over space (and time) and it can be modelled through deterministic methods. Nonetheless, there are substantial differences in the methods used to model such trend and used to interpolate values, and so, they should not be confused. Describing the nuances of these approaches is outside of the scope of this section. Yet, we advice the reader to refer to [126] for a more detailed explanation.

(b) Machine learning approaches:

From the late 20^{th} century, artificial intelligence has received increasing attention, as it can efficiently handle large amounts of data and explore complex relationships often hard to capture by the human mind. Moreover, its large adaptive capacity and automated repetitive learning make it an attractive technology in a world where data is relentlessly being generated and new knowledge produced. Machine learning is a subclass of artificial intelligence. This data-driven algorithm uses observed or historical data to identify patterns and to learn about associations between variables that can then be used to make inferences on incomplete datasets or to predict future outcomes. The processes are automated and optimized as they are provided with more input data, which allows them to iteratively improve their performance and accuracy. This 'automated repetitive learning' is what is sometimes referred to as their 'intelligence'. Depending on the level of control and supervision during the learning process we can distinguish between supervised, unsupervised and reinforced learning (Box 2.3.2.4).

Box 2.3.2.4

Supervised learning. This class of algorithms uses labelled data, i.e., input data for which the output or truth values are known, to identify the underlying patterns relating the different covariates to the output of interest. The examples used to inform the algorithm are referred to as *training dataset*. Since the output or true value is known to the modeller, the model can be validated and fine-tuned to minimize errors and improve performance, reducing bias. The quality of the training dataset is key for the model performance. A common example is RF. The most common tasks are classification, regression and forecasting.

Unsupervised learning. Here the modeller has no ground data to feed to the algorithm. Instead, the algorithm is asked to recognise correlations and relationships within the input data in order to understand and identify underlying patterns. The algorithm is then capable to organize the data based on the identified patterns. This learning method is often faster and easier to scale-up, yet there is a higher risk of bias intrinsic to its unsupervised nature. The most common tasks are clustering and dimension reduction.

Reinforcement learning. This class of learning method examines a given environment and applies trial-and-error, and delayed reward principles to identify behaviours or combinations of decisions that result in the most optimal performance for a given context. In other words, the algorithm assesses different combinations and learns from past experiences in order to adapt the response to the situation and achieve the best possible result.

RFs are a class of supervised machine learning algorithms used for classification and regression tasks. They operate as an ensemble of finite individual decision trees trained in parallel. The goal is to build a predictive model that can predict the value of a target variable using simple decision rules that have been trained on a complete dataset containing the target variable and a set of features or covariates, the *training data*. Using the features available on the training dataset, data is subsequently classified in groups with the maximum diversity possible in the target value. A decision tree will have multiple decisions rules or *nodes*. The order in which each feature is used to split the data, will be based on the one that maximizes the diversity between the groups. The prediction of each constituent tree is then combined to produce an ensemble prediction. The RF output is calculated using the mode of all its constituent trees for classification problems, and the arithmetic mean for regression problems. Provided that the constituent decision trees are uncorrelated and sufficient, the performance and accuracy of a RF will outperform that of any individual constituent tree[132]

According to Breiman et al., 2001 [132], for a RF to have a low generalization error we need the following to be met: a sufficient number of decision trees, features with good predictive capacity, and uncorrelated decision trees. The former is easily modifiable and will only be constrained by time and computational power restrictions. The second is linked to the quality of the training data and should be addressed at the collection and data processing steps. Finally, in order to achieve uncorrelated decision trees, there are two methods that are often used; bagging and feature randomness. The bagging or bootstrap aggregation method exploits the sensitivity of decision trees to changes in the training dataset. And so, instead of sub-setting the data into small chunks of data to use as training data for each tree, it uses bootstrap methods to generate a random sample with replacement from the original dataset for each tree. Random sampling with replacement refers to the fact that each value is selected at random from the original dataset and returned upon selection to the dataset, being eligible again to be selected in the same sampling iteration. Secondly, each tree in a RF is given only a subset of features selected at random from all available features, introducing feature randomness. This diversifies even more the trees, ensuring that they are not correlated [132].

Decision trees are represented by an upside down tree (Figure 2.5). The ramifications or *nodes* represent the decision rules used to split the data in *branches* or *edges*. Branches that can no longer be split by the remaining features are called *leaves*. The larger the number of internal *nodes* and *edges*, the more computational demanding the calculation is. So, the maximum number of internal nodes and edges are pre-defined a priori based on expert advice and computation restrictions. Their use in geo-sciences has increased in the last decades for they are easy to implement and transparent, facilitating its interpretation , something that artificial intelligence approaches are often criticized for. Moreover, it can handle large amounts of data with high-dimensionality, missing values, and mixed data types. Finally, it can capture non-linear relationships, ubiquitous in environmental sciences.

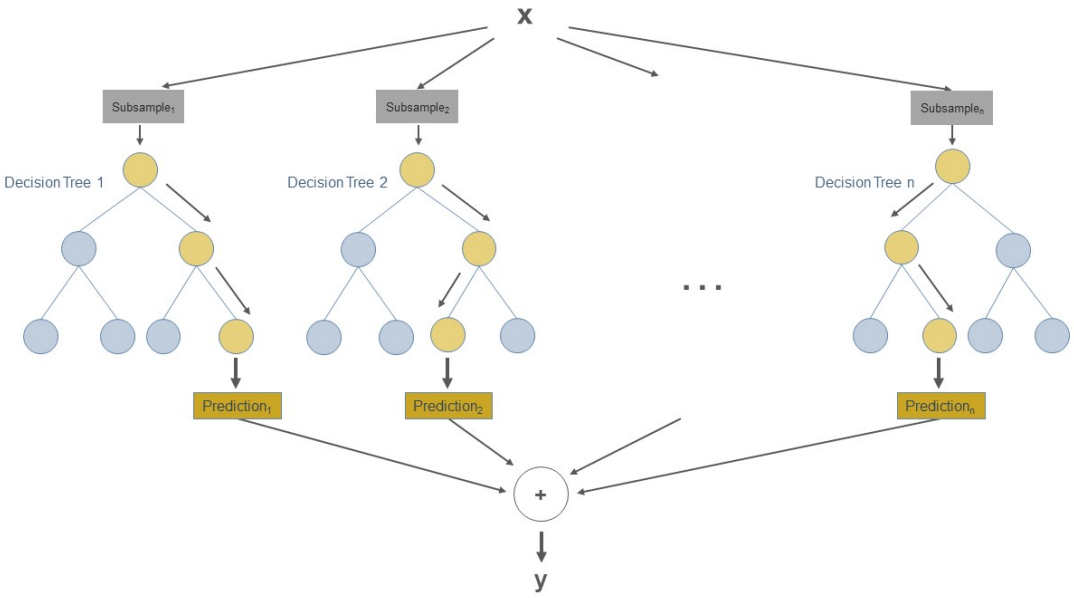


Figure 2.5: **Scheme of a Random Forest (RF).** Visualization of a RF structure

2.4 Vulnerability and exposure: determinants

As I described in section 2.1, *vulnerability* refers to the likelihood of a community to suffer from the consequences of a hazard. This disproportionate burden is a result of certain inherent physical predisposition (e.g., genetic), or related to external characteristics of the built, natural and socioeconomic environment they live in (e.g., occupation or SES). From an epidemiological point of view, we talk about *effect modification* when a particular variable has separate, different exposure effects depending on another variable which can be referred to as the *effect modifier* or *determinant*. This is to differentiate them from an interaction effect, where the joint effect of the exposure and the third variable is larger or smaller than the added independent effect of each of them [133]. The statistical approaches used are similar, yet the reporting and interpretation of the results is different. In effect modification, we are interested in the main exposure and want to evaluate whether its effects varies across population groups, e.g. gender, so that targeted interventions can be created. The most common statistical approach is the use of stratified analyses by categories of the suspect *effect modifiers*.

Compared to the large body of evidence regarding the impacts of heat and cold on health [12, 134], there is a remarkable lack of quantitative knowledge on the vulnerability of populations to temperature impacts, specially in relation to contextual determinants such as the built environment, the natural environment or socioeconomic characteristics of the area of residence [13, 135, 136]. In light of this, several international organizations, including the WHO [32] and the IPCC [1], and authors [12, 13, 137], have highlighted the importance in addressing this research gap, in particular for understudied regions, climate zones and population groups.

2.4.0.1 Individual determinants

Individual determinants include characteristics and conditions that refer to the individuals directly. For studies using secondary data, such as death certificates or hospital records recorded at the individual level , they are often collected as part of the routine procedures and provided with the data. For studies using primary data, they are collected directly from the subjects, or their first-of-kin/responsible adult, through surveys or interviews, for example. They are by far the most studied.

According to the latest literature review by Son et al., 2019 [13], the strongest evidence is for *age*; with higher risks for heat and cold mortality risk observed among the elderly — 107 out of 159 heat studies, and 47 out of 64 cold studies supporting this. Some studies have also reported a higher risk for children and infants [6], which is likely to be linked to physiological differences on their thermoregulation mechanisms as well as different behavioural and activity patterns compared to adults. For the elderly, the presence of co-morbidities, social isolation, and housing conditions seem to also be important factors [13, 138–141]. *Sex* is the second most studied individual effect modifier. In general, women have higher mortality risks associated with heat. In Son et al., 2019 [13] study, of the 74 studies assessing the role of sex on temperature health impacts, 37 found a higher risk among women, whilst 12 reported a higher risk on men and 25 no association at all. Findings remain more inconclusive for cold [13]. Several mechanisms have been suggested to explain such differences, including albeit not limited to, isolation and level of physical activity [6], occupational exposure [142], age distribution [6] and thermoregulatory and physiological mechanisms [143]. Nevertheless, the heterogeneity of results makes it hard to draw conclusions, with some studies reporting no difference or a slight higher risk for men [135, 144–146].

Education is another commonly available and explored effect modifier. However, the findings in this case are more conflicting than for age and sex. Some studies have found an increase risk with lower education both for heat [108, 141, 145–148], and cold [145]. The main train of thought is that education acts as a predictor of low SES, and thus, low educational level is likely to be associated to SES. Nevertheless, the evidence is not consistent and some studies have reported no association or even a slightly higher risk for those with higher education [141, 146]. It is hard to draw conclusions as the variable definition and its association with SES varies substantially across studies.

The evidence is more limited and less conclusive for other individual-level effect modifiers. For example, some studies have found a higher risk of heat effects for blacks whilst a lower vulnerability has been observed for Asians [149]. Noteworthy, there is no strong evidence to suggest physiological differences between *ethnic groups* [149, 150]; instead, as with educational level, ethnicity is most likely acting as a proxy for other factors such as SES, inequalities in the access to healthcare or occupational exposure. Agreeably, most of the evidence supporting ethnic (or racial) differences in the health burden of temperature - and other exposures - comes from countries which have historically suffered

from systemic inequalities, such as the USA [149, 150]. *Occupation* has also been suggested as an effect modifier in several studies [141, 142, 151]. For example, Vandentorren et al., 2006 [151] found higher risk has among manual workers as their jobs often involve physical exertion and are executed partially or in its totality outdoors. Similarly, Yang et al., 2012 [141] found higher risk among blue-collar workers in Guangzhou, China. It is important to note that occupational classification and work conditions vary substantially by setting, hindering comparisons between studies. The *location of death* has also been studied as a potential effect modifier, in particular, being hospitalized during a heat event was associated to higher risks of death in some studies [138, 152, 153] but not others [72]. People suffering from *prior medical conditions* such as obesity, Cardiovascular Diseases (CVD), high blood pressure, mental and neurological disorders , respiratory conditions or cancer have also been found to have a higher risk of health impacts from cold and heat, especially among the elderly [151]. A higher risk of death has also been reported for those *living alone* [154, 155] and *socially isolated* [17, 151]. Finally, *martial status* was also a mentioned in some studies, yet the results were inconclusive [138, 151]

2.4.0.2 Contextual determinants

Contextual determinants refer to all those characteristics not directly concerning the individuals, but instead, describing the environmental, physical, social and/or economical conditions of the areas where said individuals reside. Also referred to as *area-level, small-area, neighbourhood-level factors* or *community-level factors.*

Evidence on their role as effect modifiers is much more limited than for individual factors [13]. One of the most explored contextual determinant is SES. According to Son et al., 2019 [13] there is suggestive evidence of a higher risk of heat effects among those with a low SES. This complex indicator tries to capture the social and economic class of a community, often defined as a composite index including several indicators such as income, education, segregation or life expectancy. Its complexity and composition can look very different depending on the setting. In general, low SES has been associated to poor health status, limited access to health care and poorer living conditions, including inadequate housing and limited resources to mitigate and/or adapt to extreme temperatures [13, 156, 157]. In addition to composite indices, socioeconomic conditions can be explored through individual indicators, yet in this cases, the evidence is scarcer, and the indicators differ substantially across studies. For example, Murage et al., 2020 [17] investigated the effect modification of percentage of people who did not have English as their vehicular language, and living in social rented or owned housing. They did not find any significant effect. Silveira et al., 2021 [72] found a higher risk with areas which had the lowest average income, lowest electricity consumption and higher population density, in Rio de Janeiro, Brazil, between 2001-2018.

Building characteristics have been suggested as effect modifiers for their role in determining indoor

temperature [151, 158]. Key building characteristics in modifying the temperature exposure include: top floor flats, ventilation rate, type and quality of window glazing, geometry and fabric characteristics such as levels of insulation of the building [158]. Vandentorren et al., 2006 [151] found that during the August 2003 Heat wave (HW) in France, there was a higher risk of death among elderly living in old constructions (pre-1975; OddsRatio (OR)=1.83), on a top floor (OR=2.33), had their bedroom under the roof (OR=2.16) and had long hours of sunlight in their bedroom (OR=1.07). Greenspace has also been found to be a relevant effect modifier in several occasions [13, 17]. Multiple mechanisms have been suggested. On the one hand, the presence of greenspace reduces the local temperature and therefore exposure to heat. This is likely to have a remarkable influence on studies using coarse resolution temperature data, unique monitoring stations or assessing entire cities. This effect should be minimized in studies using spatially resolved temperature data, as the effect of greenspace on temperature is already captured by the exposure model. Nevertheless, even in those cases, some authors have reported an enduring effect [**?**] most likely through alternative mechanisms include the increase in opportunities for physical and social activities outdoors and stress reduction, which may lead to an increase in the residents' quality of life and better health, and so, reduce mortality burden attributable to non-optimum temperature.

Other contextual determinants for which there is weak evidence include: population density, heating system, healthcare facilities, proximity to water, housing quality, and air pollution level [13]. Suggestive/limited evidence has also been reported for contextual demographics, air conditioning use, urban/rural classification, climatic region and previous winter mortality [13]. Recently, some multi-city studies have attempted to shed some light to the role of contextual determinants by examining different socioeconomic city-level characteristics [159]. Sera et al., 2019 investigated 340 cities worldwide between 1985-2014 and found that cities with a high life expectancy, high Gross Domestic Product (GDP), high educational level percentage, high population density, high Gini index and high Particulate Matter of 2.5micrometers ($PM_{2.5}$) had higher heat attributable fraction [3]; whereas cities with rural or intermediate areas around, high percentage of greenspace and high hospital bed availability rates, had a lower heat attributable fraction. The effects for cold showed higher uncertainty. In another multi-city study, Bakhtsiyarava et al., 2023 [160], investigated 325 cities in South America and explored similar indicators as per Sera et al., 2019 [159]. Overall, they found stronger effects for cold-related mortality and contrasting results for heat-related mortality than those reported by [159]. In short, they found that cities with the highest income inequality had 3.45% (95% Confidence Intervals (CI): 0.33, 6.56) higher all-age cold-related excess mortality, being the only significant effect modifier for the cold-effect for the general population. High levels of poverty and segregation were associated with higher cold effects but only for the 65+ years old population. Unexpected patterns of

[3]Heat attributable fraction: proportion of deaths in the population that can be attributed to heat over all deaths.

modification by income and poverty were observed for heat-related mortality, where cities with highest levels of poverty and income inequality had smaller heat excess fractions than their counter-partners. The authors suggest that the differences with Sera et al., 2019 [159] could be attributed to the use of tertiles instead of continuous variables to allow for non-linear effects, or due to innate difference in the cities and time period explored, just to mention some. This, once again, supports the need for city-specific and intra-city studies.

2.5 Scale matters: the special case of urban settings

2.5.1 City-level studies

Studies investigating temperature specifically in urban areas are scarce. Some studies have investigated large regions within which there are indeed cities, yet model evaluation has not been conducted locally for specific cities, or at least not reported as such, hindering comparison between urban settings [95, 161]. Others have used larger unit of analyses, such as province[80, 162], region [69] or country [69, 70]. Although they provide important insights on the potential role of effect modifiers, the ecological fallacy error makes them inadequate to directly extrapolate its results to smaller regions.

Cities in particular behave remarkably different than rural areas and peri-urban areas when it comes to temperature effects [15]. The physical characteristics of urban areas, such as building density and layout, building material or prevalence and distribution of green and blue spaces, significantly modify the energy balance and wind patterns near the surface, making cities a unique setting when it comes to temperature distribution [163–165]. This results in drastic temperature changes within the scale of just a few hundred meters[166]. Moreover, the resilience and adaptation infrastructure and technology available to cities is substantially different to other areas. In recognition to the uniqueness of cities, multi-city studies have been on the rise in recent years[36]. By collecting city-level data across multiple settings they can achieve higher statistical power and higher contrast of variables, especially if they compare large and diverse regions, sometimes extending across the globe. Moreover, city-wide data on effect modifiers, temperature exposure and health outcome is often more readily available than more granular data (e.g., intra-city data), making it easier to access. Multi-city studies have provided a unique insight into the differential behaviour of cities as opposed to rural and peri-urban areas. Currently, there are two large multi-city networks with interest in quantifying and deepening our knowledge on the health burden of non-optimal temperature: the Multi-City Consortium (MCC) Collaborative Research Network and *Salud Urbana en América Latina* - Urban Health in Latin America (SALURBAL).

The MCC Collaborative Research Network[167] is an international partnership of research teams set up to generate epidemiological evidence on the connections between environmental stressors, cli-

mate, and health. Among the environmental stressors most investigated by the network there is non-optimal temperatures and temperature extremes, with several publications [30, 76, 159, 168] . To date, the network holds the most extensive data on environmental exposures, health outcomes, and climate projections from numerous cities across various countries worldwide. This dataset enables standardized analyses of local data to address specific research inquiries concerning global associations between the environment and health [169]. The network is coordinated by a team at the London School of Hygiene and Tropical Medicine, with the guidance of a Scientific Committee that supervises and directs the research efforts. To date, the MCC dataset includes 18 cities in Brazil, one of which is São Paulo, with data between 1997 and 2011 [3].

The SALURBAL [170] network was set up in April 2017 . Coordinated by Drexel University and funded by the Wellcome Trust, this network includes 14 institutional partners across the USA and Latin America, including the University of São Paulo in Brazil. Their aims are (i) to investigate the drivers of health in cities, (ii) to evaluate the impact of urban policies and interventions on health and the environment, (iii) to apply systems thinking to analyse the relationship between the urban environment, health and environmental impact, and (iv) to collaborate and co-create with policymakers and media to disseminate the research and ensure its integration in policy and action. Among the environmental hazards explored there is the health burden of non-optimal and extreme temperatures, with several publications [160, 171, 172]. This network represents a pioneering endeavor in consolidating an extensive compilation of environmental factors, health-related variables, and socioeconomic data, thereby setting a remarkable precedent. The network has collected, harmonized and standardized data from over 300 cities in 11 countries, laying the foundations for conducting standardized analyses on local conditions to address concrete research questions. This network represents a significant breakthrough for Latin America, a region that currently has 80% of its population living in urban areas and that bears a pressing need for comprehensive research in this domain.

2.5.2 Intra-city studies

In this chapter, I use the term *intra-city studies* to refer to studies which investigate whether the temperature-mortality association differs within a city, focusing on differences between areas and groups inside a city. Slowly, *intra-city studies* are becoming a feasible objective as more granular data becomes available, both for individual and contextual determinants, and temperature data (refer to section 2.3.2). Nevertheless, the evidence remains scarce. Among the available studies, one can distinguish two types depending on whether their objective is (i) to characterise small scale spatial patterns or (ii) to purely investigate the temperature-health association by different contextual characteristics. Both for (i) and (ii) the statistical approaches most commonly used are, as for larger regions, the case-crossover design [19, 70, 173] or the time-series design [16, 174, 175].

For (i) the focus is on spatial patterns at a high spatial granularity which permits, among other things, identification of the so-called *hotspots* or areas with a higher than expected risk. These are of interest to policy-makers and urban planers as they can pin-point where in the city an intervention is needed. A major challenge when performing small-area analyses is the availability of high spatial resolution data and the instability of the small-area specific estimates, which creates a lot of noise. Thus, most spatial assessments studies have predominantly utilized aggregated data at a macro-level geographical scale, leading to limitations in capturing fine-grained spatial nuances and identifying area-specific vulnerabilities [159, 176]. Only a few studies have employed finer geographic scales, mostly focusing in urban areas [17–19, 175]. For example, Xu et al., 2013 [19] conducted an analysis in Barcelona, focusing on census tracts and contrasting mortality rates on "hot" days versus other days during the warm seasons from 1999 to 2006. They observed a significant 30% increase in all-cause mortality on "hot" days and identified certain factors, such as the percentage of old buildings, manual workers, and residents with limited access to green spaces, as effect modifiers. Only a handful explored larger regions [16, 173, 174, 177]. For example, Bennett et al., 2014 [173] applied Bayesian spatial methods to the district-specific CVD mortality risks to borrow information from neighbouring districts across England and Wales (2001-2010); the extent of the influence was determined by the variance in the estimated effects in each district and the empirical similarity with their neighbours. The risk estimates had been calculated using a case-cross over design. Recently, Gasparrini et al., 2022 [16] have proposed a new approach based on an extension of the two-stage DLNM framework for case time series design using nested spatial units and adding a downscaling procedure. The case time series design defines case-specific series of continuous sequential observations (i.e. multiple time-series), in this case, area-specific time series. A full description of the method can be found in [16, 83]. As a proof of concept, the authors estimated the temperature-mortality risk for 34,753 Lower Super Output Areas (LSOA) within 348 Local Authority District (LAD)s across England and Wales between 2000 and 2019. In the first stage, they estimated the temperature-mortality association for each LAD using multiple LSOA-specific series modelled through a conditional Poisson regression. This step was undertaken for four age groups separately. In the second stage, the risks were accumulated over the lag dimension and pooled across age groups using using a multivariate repeated-measure meta-regression. At this stage, the authors included LAD-level information on potential effect modifiers, such as SES indicator. The last step consists on the downscaling of the risks from LAD to the LSOA, using the coefficients of the effect modifiers used in the meta-regression. It is important to note that the indicators must be available at the LAD and LSOA level for this approach to work. In short, this approach allows to borrow information across different spatial units to leverage the uncertainty in the estimates. Noteworthy, it builds on the assumption that the influence of the effect modifiers on the temperature-mortality association is identical across spatial units, which may not always fully hold

specially if the spatial units are substantially different in size.

For (ii) the goal is to explore the differential risk across groups which can be defined by individual or contextual determinants. Here, despite the mortality and effect modifier data is also obtained at a small-area level as in (i), the association is modelled jointly for all the spatial units that belong to a given category of an effect modifier (e.g., all census tracts belonging to the most deprived quintile of the city SES indicator). In short, the exposure and effect modifier data assimilation is done at the small-area level, yet the analyses are performed by category of effect modifier, in a stratified analyses. This reduces the uncertainty, giving more robust estimates. Limitations include the lack of information on the spatial distribution of the risk over the city and the fact that only one (or a few, depending on statistical power) effect modifiers can be analyzed at the time. This contrasts with the approaches in (i) where multivariate analyses can be performed. Moreover, most of the studies investigating effect modifiers within a city have focused on individual determinants [141, 178] with only a few describing contextual determinants [17, 179]. Of these, most use temperature data from a unique meteorological station, an average of available stations in the city or a coarse model or reanalyses dataset. Only a few studies have focused on examining the distinct patterns in temperature-health outcomes that exists within a city [16, 17, 67, 173]. Saucy et al., 2021 [67] examined the effect temperature on acute cardiovascular mortality in Zurich, Switzerland between 2000-2015. In their analysis, they used $2km^2$ temperature models provided by MeteoSwiss, the national Swiss meteorological services. Murage et al., 2020 [17] investigated the role of contextual and individual factors in the temperature-mortality association in London using the MEDMI modelled data from the UK Meteorological Office, available at a resolution of $500m^2$.

In general, some other limitations of intra-city studies include: (i) use of temperature data from nearest or referent meteorological station[3, 10, 41, 110, 178, 179], average of all stations [71, 72] or coarse resolution temperature datasets [36, 70], raising concerns about exposure miss-classification bias in their findings; (ii) some are restricted to individual determinants as they do not have access to spatially granular data and so, they can not link the data to contextual determinants[141, 178], and (iii) they often suffer from low statistical power as the mortality counts are inevitably lower than if the study area was larger.

2.6 Public health strategies and temperature

Weather and climate, which include temperature, have a significant impact on the health. On preparing, responding and adapting to this indisputable risk depend the livelihoods of communities. Individual efforts can help mitigate and adapt to these increasing risks, such as using household air conditioning or changing certain behavioral practises. However, most of them require of a certain

level of organization, governance and leadership. Thus, developing evidence and climate informed strategies, has been recognized as an essential tool to prepare, respond and adapt to the increasing climate-sensitive disease risk and outbreaks.

2.6.1 Health and climate change plans or strategies

National health and climate change plans or strategies consider the risks associated to climate change and define different adaptation and resilient priorities and strategies to achieve them. They often involve a multi-sector response with a wide variety of actions and agents. For example, they may include investing in urban developments for climate mitigation, such as the instalment of green roofs and vertical gardens to achieve local cooling, or they may involve target communication of risk to vulnerable groups to advice of certain behaviours to lower their risk. The climate-sensitive health risks affecting each country is different, as well as the financial resources, leadership and governance structures, and so, these national plans vary substantially.

2.6.2 Climate-informed health surveillance and early warning systems

Climate-informed health surveillance and early warning systems are a key element of these national strategies. These systems aim at detecting, investigating and responding to public health threats in a timely and systematic manner in order to minimize the impacts. Heat Health Warning System (HHWS) are weather-based alerts activated upon reaching certain heat levels that are considered damaging for the health of a community. Heat-Health Action Plans (HHAP)s are the evidence-based comprehensive strategies activated by HHWS and designed to minimize the impacts of hot weather extremes on health. They have repeatedly been shown to be effective in reducing heat-related morbidity and mortality [180, 181], and economic impacts [182]. In their 2014 report, the IPCC highlighted the need of HHWS and HHAP to effectively reduce the health burden of heat [1]. Currently, the development and improvement of HHWS is among the priorities of the World Meteorological Organization (WMO), and the WHO.

HHWS's activation is defined by a certain thresholds of heat tolerance to extreme weather by the population of interest. According to the WHO recommendations [32], these thresholds should be based on epidemiological evidence of the heat-health risks. Not only that but, ideally, it should reflect the specific risk of the population for which the HHWS is being designed, as we know that heat-health risk profiles vary substantially depending on the time period, region and underlying vulnerability profiles, among other factors. Several HHWS are not based on epidemiological studies and those which are, often do not have access to epidemiological evidence specific to the area for which the HHWS is being developed [32, 183]. Bulkeley et al., 2010 [184] argue that the lack of data and expertise at the local level is the most critical obstacle to policy development when it comes to adaptation. Finally, almost

all HHWS use temporally static thresholds for simplicity [32, 181, 183]. Nonetheless, heat-health risk can vary substantially over time and by season. Temporally flexible HHWS have been successfully implemented in some settings, such as Spain [183] and Quebec, Canada [185], yet they are not common.

LMIC are particularly vulnerable to the effects of weather extremes and so, they would benefit enormously of the implementation of HHWS, particularly in areas that are highly populated [186]. Unfortunately, HHWS are primarily in use in high-income countries in Europe [181, 183] and North America [185, 187], and in some parts of Australia and Asia [188, 189]. Needless to say, findings from high-income countries can seldom translate to LMIC, which in addition to belonging to a different climate region, have different social, economic and political structures, with often less access to infrastructure, technology and services, all key factors in defining the health burden of temperature[12].

Regional spatial variability in the heat-health risk is in some measure well-accepted and can be seen incorporated in most HHWS. For example, McLean et al., 2018 [187], suggested specific threshold for British Columbia based on epidemiological evidence specific for each of its four regions. Thresholds varied as much as 5°C across regions, highlighting the differences in risk, and the inappropriateness of a unique threshold. In Europe, most HHWS are at a sub-national scale, ranging from regional plans (e.g., Austria, France, Switzerland, Netherlands) or smaller units (e.g, Belgium and Spain have province-specific plans, Sweden and Germany have county level plans) [183] . Out of the 16 HHWS active in Europe, only Lisbon in Portugal and several Italian cities, including Rome have a specific plan [183]. When available, these city-specific plans have been shown to better protect the population over regional or national approaches, which highlights the value of city-specific thresholds. Unfortunately, spatial variability within cities is understudied, and its implementation in HHWS neglected.

The WHO health and climate change survey is a triennial investigation of the progress of governments in addressing the health risks of climate change among the 194 WHO Member States and a small number of non-Member Territories. In the 2021 report, out of the 95 respondent countries, 49 (52%) indicated having a health and climate change strategic plan in place, and 24 under development (25%). Almost half of the countries with a strategy or plan, indicated that insufficient research and evidence was among the main barriers to implementation. And of the countries with national strategies in place, a very small number mentioned investigating and addressing vulnerable groups (i.e., 11 studies mentioned gender differences, 17 age differences, and only 8 identified other vulnerable groups)[190]. Of the 31 national plans that were submitted to the revision board, only 15 of them considered heat-related illnesses. Regarding surveillance systems, only 42% of countries reported having a surveillance system for injury and mortality from extreme weather events and 20% for heat-related illnesses. Finally, only 39% of countries had a HHWS and, of these, only 40-75% had been evaluated (percentage range reflects different evaluation components). Brazil was among the countries investigated. The country's first national health and climate change plan/strategy was launched in May 2016 [190, 191].

However, minister respondents acknowledged that the level of implementation is moderate, with only some of the plan/strategy priorities being addressed. The country also has a health surveillance system in place for airborne and respiratory illnesses, vector and waterborne diseases, malnutrition and food-borne diseases, mental health and psychological illnesses, non-communicable diseases and injury and mortality from extreme weather events. Yet, only the surveillance system for airborne and respiratory illnesses, and injury and mortality from extreme weather events are informed by meteorological data. Finally, Brazil has an warning system for landslide, flooding events and droughts operating in 958 out of 5,568 municipalities[192], yet to date, there is no climate-informed HHWS or HHAP.

In short, health and climate change research are a crucial component of early warning systems and health and climate change action plans as they improve our understanding of the current and future distribution of the hazard, health hazards and vulnerability determinants. The large heterogeneity in the spatiotemporal distribution of the health burden of temperature across regions, cities and time periods supports the importance of performing location-specific and periodic epidemiological studies to provide accurate and suitable information to feed into HHWSs and HHAPs. This is specially of interest within cities, where the weather conditions as well as the social, economic and demographic characteristics of the population show a high level of heterogeneity in small temporal and spatial scales. In this context, evidence such as that presented in this thesis can be valuable to evaluate whether there is a need for intra-city specific HHWS and to define the thresholds for such a system in the context of São Paulo.

2.7 São Paulo, Brazil

Latin America is one of the most urbanized regions in the globe, with around 85% of population already living in cities [22] and therefore, has a large population at risk of urban heat exposure [171]. Located in the south-eastern part of Brazil, and with over 22 million inhabitants, São Paulo metropolitan area, also known as Metropolitan Area of São Paulo (MASP), is the largest urban conglomeration in South America, hosting 10% of the total population in Brazil, and the fourth most populated metropolis in the world [22]. Its population is projected to reach nearly 24 million by 2030 [22]. More than half of the population, around 11.7 million, live in 1 of the 36 other municipalities included in the metropolitan area, the municipality of São Paulo. The municipality expands over an area of 1,521 km^2, and it is divided in 32 sub-prefectures, 96 districts and 18,363 census tracts. Similar to many megacities in less developed regions, São Paulo concentrates both the resources and problems that plague urban systems such as ingrained social inequities, high levels of air pollution, lack of greenspaces, UHI or high unemployment rates among others. Thus, São Paulo is suitable to investigate the complex characteristics of climate change impacts, including those from temperature, the focus in this project.

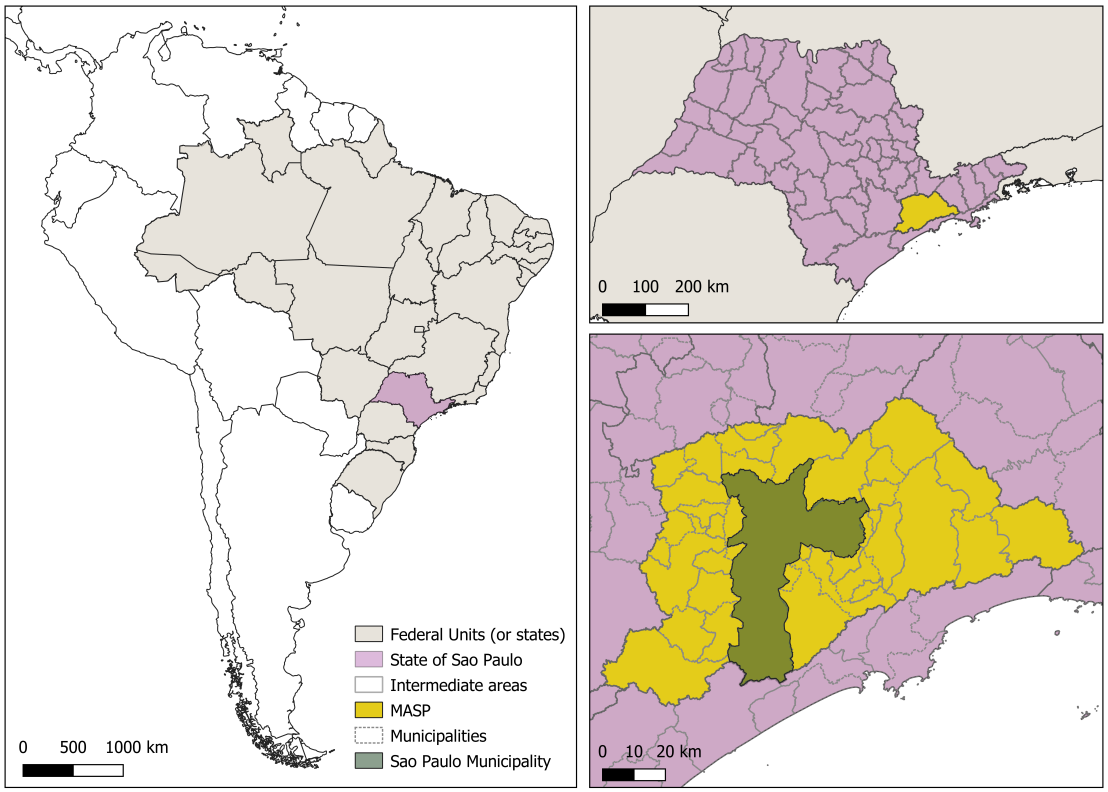


Figure 2.6: **Map of São Paulo**. Map displaying the state (purple), the metropolitan area or MASP (yellow), and the municipality of São Paulo (green).

2.7.1 Temperature: past, present and future

The municipality of São Paulo is located in the southern hemisphere, close to the Tropic of Capricorn (23° 33S, 46° 39 W), 80 km west from the Atlantic Ocean, on a plateau at an altitude between 700 and 1,000m above mean sea level. The ocean and land breezes and topographic features of the area are key factors influencing the climate. According to the Köppen-Geiger classification, it has a temperate climate without dry season and with hot summers (Cfa, previously referred as Humid subtropical climate [193]) in the North, and with warm summers (Cfb, previously referred as Temperate oceanic climate) in the South. Temperature shows a clear seasonality with the hottest period being December to March (mean temperature of 24°C) and the coldest period being May to August (mean temperature 20°C) (Figure 2.7). The maximum and minimum temperatures follow the same seasonality pattern with an oscillation of +6 or -6°C in reference to the mean, respectively. During the cold months, precipitation drops down to below 80mm per month. Precipitations are frequent during the warm months registering in average over 160mm monthly.

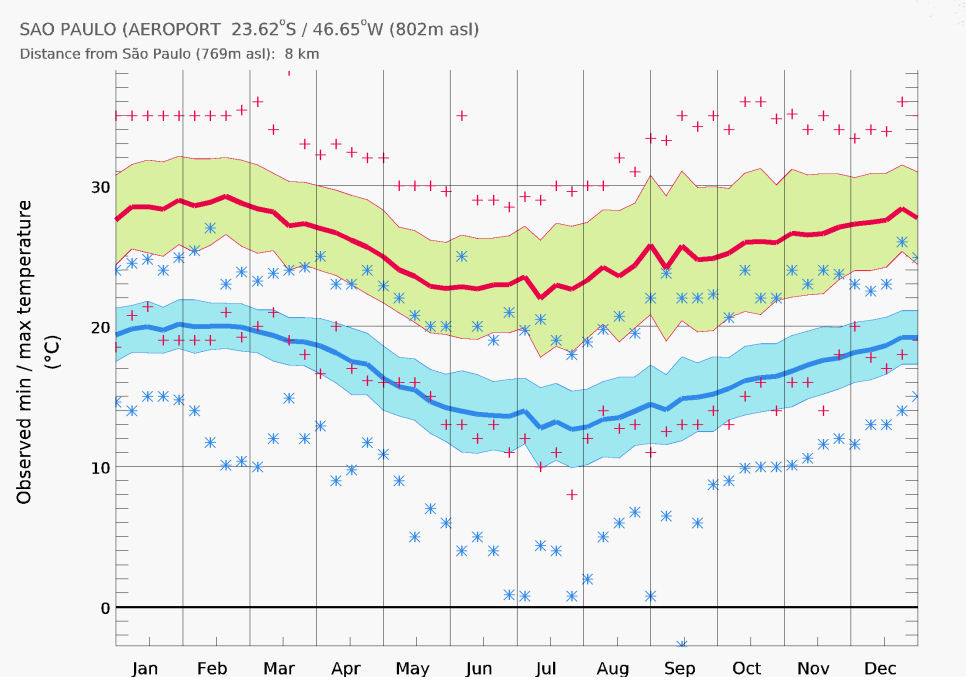


Figure 2.7: **Maximum and minimum temperature in São Paulo.** Average maximum (solid red line) and minimum temperature (solid blue line) for São Paulo. Extreme values that fell outside the usual range are represented by + and * characters for heat and cold, respectively. Based on 10 years of measurements from the Airport Meteorological station. Source: MeteoBlue[194]

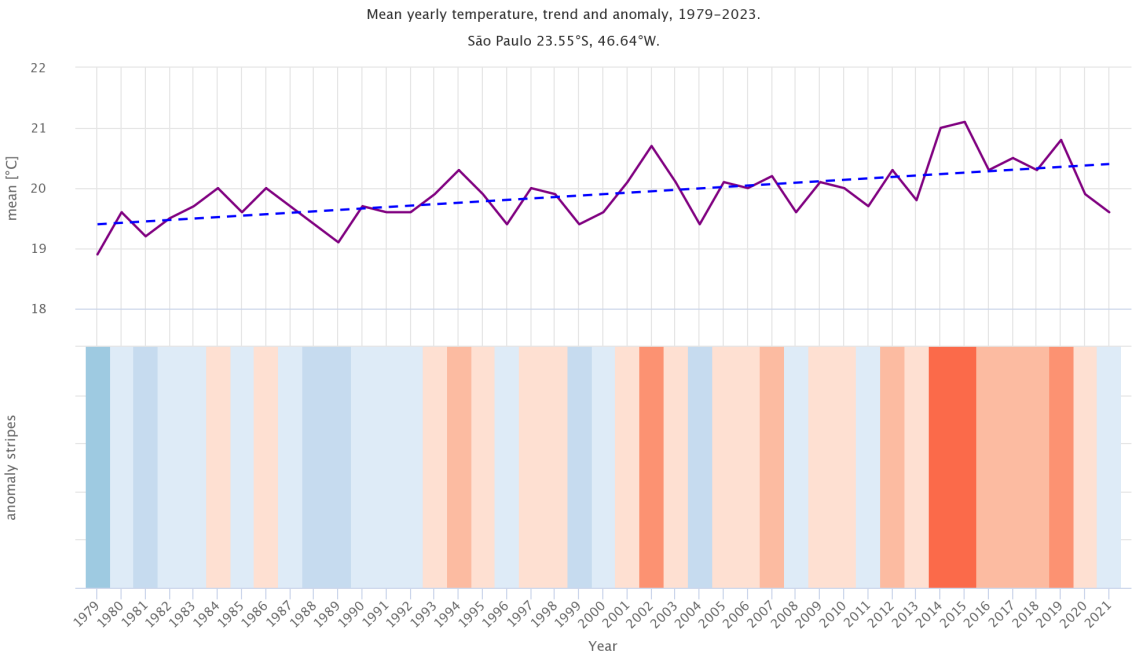
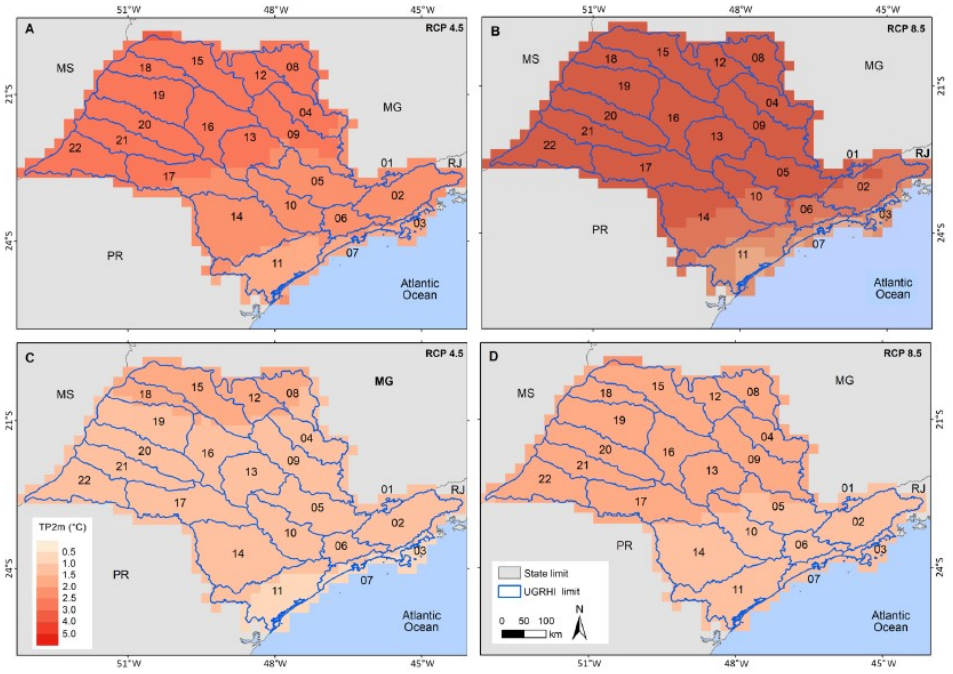


Figure 2.8: **Annual mean temperature for São Paulo between 1980 to 2020.**(Top) Annual mean temperature (solid red line) and linear trend (dashed blue line). (Bottom) Warming stripes plot where each stripe represents the mean temperature for that year relative to the average over the period 1980-2020. Colored as a gradient from coldest (blue) to warmest (red) temperatures. Temperature data from ECMWF ERA5 dataset. Source: MeteoBlue [194]

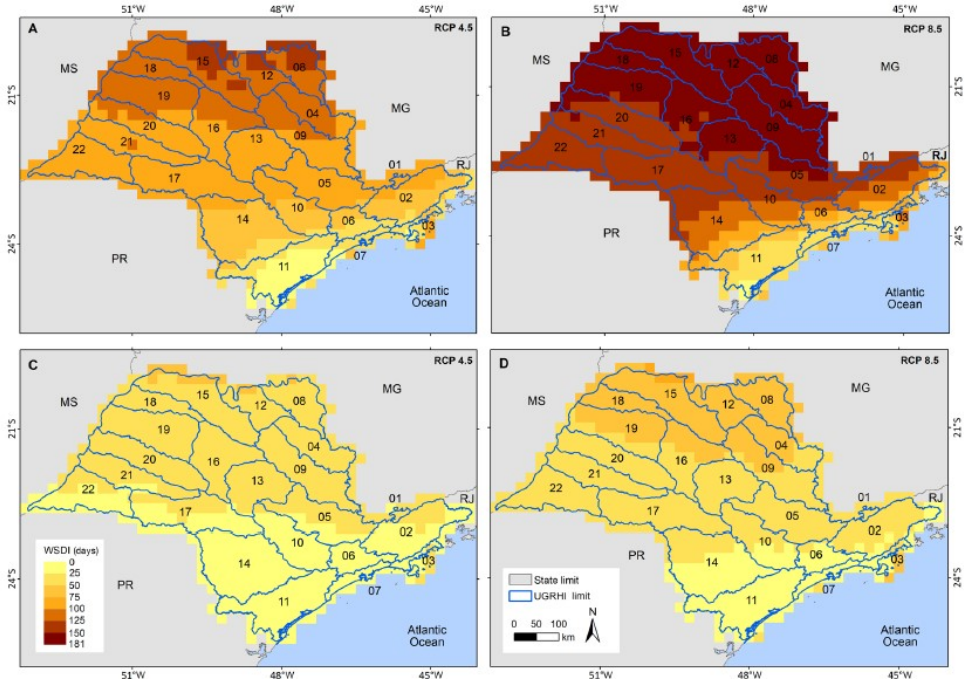
Over the years, there has been an increase in the annual mean temperature (Figure 2.8), which has been primarily linked to both climate change and the rapid urban growth experienced by the area in the last 30 years [24]. Based on ERA5 2m ambient temperature retrieves between 1980 and 2020, the annual mean temperature has increased about 1.0°C (Figure 2.8 (top)). Overall, there was a progressive increase in the deviance towards higher temperatures, with the hottest years being 2002, 2014, 2015 ad 2019, coinciding with el Niño years(Figure 2.8). An increase in the number and duration of extreme temperature events has also been described in the literature. Geirinhas et al., [195] investigated Manaus, Receife and São Paulo, all cities in Brazil and found that São Paulo had the largest number of heatwave days, with the largest number in 2014 when 50 days were registered.

São Paulo, as most cities, has a large proportion of land devoid of green spaces, high population density, high concentration of high-rising buildings, large concentrations of air pollution and large prevalence of heat-absorbing construction materials, all contributing to the UHI. Research from São Paulo indicates a variation of more than 10°C in surface temperature between the city center and the surrounding rural and peri-urban areas [196]. In São Paulo, the UHI has its maximum intensity during the day, between noon and late afternoon, reaching its peak at around 17h [24, 196]. Narciso de Lima et al., 2018 [24] investigated the impacts of urbanization on the local climate in São Paulo and showed that since the mid-70s, São Paulo temperature has increased steadily with changes in land use cover, primarily in relation to reduction in greenspaces. These changes in temperature affect both the average temperature with an increase of approximately 1.1°C every 20 years in the mean annual maximum temperature, and the frequency of extreme temperature events[24] .

Recently, Armani et al., 2022 [197] projected the climate for the State of São Paulo considering the Representative Concentration Pathway (RCP) 4.5 and 8.5 scenarios. Figure 2.9a shows the deviance in the maximum (top) and minimum (bottom) annual mean temperature projected for the period 2020-2050 in reference to the historical period 1961-1990 for the RCP4.5 (left) and RCP8.5(right). In both scenarios, they predicted a substantial increase in both minimum and maximum temperature, with a slightly higher deviance for the RCP8.5 scenario (RCP8.5: 0.5-1.0°C minimum , 2.0-2.5°C maximum; RCP8.5: 0.5-1.5°C minimum; 2.5-3.0°C maximum)(Figure 2.9a). These are in line with other published projections for the region [198]. In the paper, they also investigate the deviations in annual maximum and minimum temperature. For minimum annual temperature the deviance were similar to the mean annual temperature. For the maximum annual temperature, the authors report maximum deviance of up to 4.0 for the RCP4.5 scenario and 5.0°C for the RCP8.5 scenario. Finally, the number of consecutive days in the year with a maximum temperature above the 90^{th} percentile was also projected to increase between 25 and 150 days for the RCP4.5 scenario (Figure 2.9b (right)) and up to 181 days in the RCP8.5 (Figure 2.9b (left)). In all projections, the North East of the state suffered from the largest deviations; São Paulo had a mild impact relative to the rest of municipalities.



(a) Projected deviance in annual mean temperature



(b) Projected deviance in number of consecutive dates with extreme heat

Figure 2.9: **Temperature projections for the State of São Paulo under RCP4.5 and 8.5** Maximum (top) and minimum (bottom) deviance in (a) annual mean temperature and (b) number of consecutive days within a year with a maximum temperature above the 90^{th} percentile projected for 2020-2050 in reference to the historical period 1961-1990, for the RCP4.5 (left) and RCP8.5 scenarios (right). São Paulo municipality shown as number "06". Source: Armani et al., 2022 [197]

Finally, HW are expected to increase globally and also in South America. Feron et al. 2019 [199] investigated the duration, frequency and number of HW per hot season in some of South America's largest cities. For São Paulo they projected a 2-fold increase in duration (from 8 to 18 days), an 8-fold increase in the frequency (from 3 to 24 days) and a 7-fold increase in the number of HW (from 0.4 to

3) for the period 2046-2055 in reference to the period 1961-1990. Finally, they project an increase in the percentage of days during the hot season exceeding the 95^{th} percentile of the baseline distribution of maximum temperature from 5% to an appalling 27% under RCP4.5 and of 40% under RCP8.5.

2.7.2 Individual and contextual determinants

A large population density, inequalities, rapid and irregular urban growth are just some of the challenges that add to the climate risk, aggravating its impacts. Over the years, a combination of lack of employment and low salaries, has brought a rapid and disordered occupation of the land, causing urban sprawl and the rise of informal settlements, also known as *favelas* [200]. In 2010, about 15% of the census tracts were classified as sub-optimal or precarious settlements[201]. People living in these settlements are often of low SES, live in poorer conditions and are more vulnerable to suffer from poorer health. The city's current Master Plan [202] has at its core, addressing such irregular growth and reducing the inequalities that derive from it. São Paulo's master plan is a pioneer and new approach to urban planning, putting the addressing of inequalities at its core. One key distinctive characteristic is that the new plan recognized and registered slum-dwellers as residents. By doing so, their rights are recognized and have to be protected. This includes basic needs such as access to water and electricity. This is not the case in many other cities worldwide where slum-dwellers live outside of the law and are not recognized as citizens, making it really easy for governments and developers to neglect their needs, as there is no legal consequence. Despite the ingrained inequalities and housing conditions affecting São Paulo, this city is also Brazil's technological, financial, research and industrial hub. This gives it an advantageous position and the potential to lead the transformative process required to address climate change [191].

The elderly population is projected to increase globally and São Paulo is no exception. According to [203], by the end of the century, the elderly population will increase 142% by 2050 and 175% by 2100 in relation to 2020.

2.7.3 Climate adaptation policy in São Paulo

São Paulo is the leading state in Brazil when it comes to climate adaptation strategies [191, 204]. Two events marked São Paulo's journey to climate change action. The first one in 2003 when the city joined the Cities for Climate Protection Campaign[205]; the second one, in 2005, when it joined the Cities Climate Leadership Group (C40) [206]. Eventually, this lead to the launch, in 2009, of São Paulo's Municipal Climate Change Policy (Law No. 13.798/2009 and Decree No. 55.947/2010), becoming the first city in Brazil to set specific mandatory green house gas emission reductions. An advisory body was created to support its implementation, engaging a wide range of stakeholders (e.g., municipal government, civil society). In 2015, the State of São Paulo enacted the Climate Protocol,

an optional endeavor urging the private sector to collaborate in tackling climate change. In July 2021, São Paulo formally adhered to the Race to Zero [207] and Race to Resilience [208] campaigns through the Decree No. 65.881 [209]. As a result, the state of São Paulo published a first draft of the "Strategic Guidelines and Actions for the Climate Action Plan of the State of São Paulo - Net Zero 2050", in which they defined the state's climate aspirations for the next three decades [210, 211]. Among these aspirations, they mention briefly the need to "invest, on an ongoing basis, in improving the issuance of early warnings through the technical-scientific development of the monitoring of risk areas" including among other climate hazards, heat [211]. The decree includes provisions for developing a Climate Adaptation Plan, which will take into account the assessment of risks and vulnerabilities related to potential climate events linked to global warming. Despite all these efforts to integrate climate change action into policies, São Paulo has lagged behind in results. Lack of collaboration across stakeholders, presence of conflicting interests, and insufficient and inadequate data and knowledge, monitoring and development of preventive actions [191], have been suggested as reasons for this failure.

Finally, a part from the climate-specific policies, there are other initiatives that address some climate issues indirectly. For example, NGOs, civil society and stakeholders have successfully implemented some initiatives. One example is the *Refresca São Paulo* initiative which aims at adapting slums to heat island stress and water insecurity by introducing green areas in the city [212]. Another example is the new city's Master Plan [213]. This comprehensive urban development strategy outlines the guidelines and policies for the city's growth, including land use, transportation, housing, environmental protection, and social inclusion. It builds on a progressive urban planning agenda that has helped dramatically reduce inequalities in São Paulo over the past 15 years. The latest Master Plan was launched in 2014 (*Plano Diretor Estrategico - PDE , 2014*) [202] and it is remarkably consistent with the international best practices and Sustainable Development Goals. It has been recognized by the UN as an exemplary and ambitious plan for sustainable and equitable urban development. Despite the Master Plan does not include any explicit mention of climate change, several of its priority areas have clear synergies with climate change mitigation and adaptation, such as green infrastructure, reduction of inequalities or prioritizing non-motorized transport.

2.7.4 Heath and heat strategies

São Paulo has a city-level Cold-Health Action Plan (CHAP) managed by the *Centro de Gerenciamento de emergencias climaticas da Prefeitura de São Paulo* [214] aimed at protecting the homeless population. This program is operational from May. Two levels of warning exit, a first level issued when temperatures reach values below 13°C, and a second level activated if temperatures drop below 10. Upon exceeding one of the above-mentioned temperature thresholds, the *Coordenadoria Municipal de Defesa Civil* and the the *Secretaria Muncipal de Assistencia e Desenvolvimento Social* are informed

and they activate the following : increase the capacity of homeless shelters to match the demand; extension of the day-care shelters to cover the coldest hours (16h to 9h) every day of the week, and mobilization of social services teams to attend the needs of homeless individuals on site and direct them to temporary shelters. To date, São Paulo does not have a HHWS for heat nor does it have a HHAP in place.

2.7.5 Temperature-mortality epidemiological evidence

Overall temperature-mortality association. A number of studies have investigated the temperature-mortality association in São Paulo, either focusing only on the city [108, 179], comparing a small number of cities[215], or as part of a multi-national multi-city study [30, 53, 171]. Table 2.1 shows these studies and indicates the time period and outcome investigated. In general, despite covering different time periods, all studies described the temperature-mortality association as a U-shaped relationship with a long and moderate tail for cold effects and a short yet high tail for heat. Hajat et al., 2005 [53] reported interesting information about the harvest effect in São Paulo useful to inform the maximum lag of local models. Using data between 1991 and 1994 the authors showed that an intermediate harvesting effect was present, extending over 3 weeks, meaning that increases in mortality at high temperatures were not only due to abrupt deaths but also represent substantial shortening of life of individuals whose health was already compromised. This effects were stronger in Delhi, India where there was a higher prevalence of deaths for infectious diseases. This findings show that populations in LMICs where life-threatening infections are common, may suffer from a stronger harvesting effect, and suggest the use of longer lags to account for the harvesting effect.

Vulnerability and exposure. With regards to vulnerable populations, most studies available for São Paulo explore individual determinants [53, 179, 215], primarily investigating differences in age, sex and educational attainment. In general, they describe higher risk of heat-related mortality for the elderly, women and those with low educational attainment. Studies exploring other individual-level characteristics such as occupation, ethnicity or martial status are lacking. Regarding contextual determinants, all the available evidence for São Paulo comes from multi-city studies [159, 160] using cities as their unit of analyses reporting conflicting results (refer to Section 2.4 for more detail). To my knowledge, only one study has investigated contextual determinants within the city of São Paulo. In this study, Gouveia and colleagues investigated the role of SES on the temperature-mortality association between 1990-94 by age groups and found no significant association. As it is common in city-level studies available to date, the exposure assessment was based on a one unique weather station, missing the spatial variability of the exposure. To my knowledge, there is yet no study exploring contextual determinants within the city of São Paulo using high-resolution temperature data and so, it remains a research gap worth investigating.

Future projections. In a global multi-city study in 2017, Gasparrini and colleagues [198] projected the heat and cold-related mortality under multiple climate scenarios (RCP2.6, 4.5, 6.0 and 8.5) for the period 2050-59 and 2090-99. Overall, they observed a substantial decrease in the cold-related mortality risk across all scenarios, which counterbalanced the increase in heat-related mortality for the 2050-59 period. However, by the end of the 21^{st} century, the heat-related mortality increase was significantly larger than the reduction in cold-related deaths, causing a substantial increase in the net mortality associated to non-optimal temperatures across all scenarios, but particularly alarming for RCP8.5. For Brazil, the net increase under the worst scenario would raise from +1.0(-3.8;5.2) in 2050-59 to +5.1(-20.1;-21.2) in 2090-99. The change in both heat and cold excess mortality showed a high heterogeneity by region and city. India and the North of Brazil registered the largest net increase, between 15-50% by 2090-99 under the worst scenario. For São Paulo the net change was between 0 and 5%. These findings debunk the believe that heat-related mortality will be counterbalanced by a reduction in cold-related mortality, stress the urgency of mitigating and adapting to future temperature risk, and once again, show that risk profiles are highly region and city-specific.

Regarding extreme temperatures, Diniz et al., 2020 [203] projected the excess mortality associated to extreme temperature events in the metropolitan region of São Paulo between 2030-50 and 2079-99. The first used a quasi-Poisson regression with DLNM to assess the heatwave-related excess mortality among the elderly historically (1996–2016). They defined a heat wave as a period of at least three consecutive days with maximum temperatures above the 90^{th} percentile of the maximum daily temperature of the reference period (1996–2016), accounting for day of the week, seasonality and long term trends. Using temperature projections for 2030-50 and 2079-99 under RCP4.5 and RCP8.5 they then estimated the excess mortality each scenario assuming 'no-adaptation' and 'adaptation'. In this study, adaptation was defined as the capacity of the population to tolerate higher temperatures and so, pragmatically considered by re-calculating the HW with future temperature distributions. Under the adaptation assumption, the duration and frequency remained rather similar for both scenarios and time periods investigated. However, the intensity increased substantially for all scenarios and time periods. For the no-adaptation assumption, the authors reported a dramatic increases in all HW characteristics. Under the adaptation assumption, the authors projected an increase in cardiovascular and respiratory diseases among females for all scenarios and time periods studied, with the largest HW-related excess mortality per year for CVD under the RCP8.5 scenario in 2079-99 (78 excess deaths per year/100,000). The projected estimates for males also progressively increased with worsening scenario and larger time period for CVD, but not for respiratory diseases. For the non-adaptation assumption the estimates doubled those of the adaptation assumption. In this case, the differences between scenarios were more significant, with the RCP8.5 scenario showing the largest excess mortality estimates mainly from cardiovascular diseases in women (587 deaths/100,000 inhabitants per year).

Table 2.1: **Epidemiological studies on health impacts of temperature in São Paulo.** Selection of epidemiological studies investigating the temperature effects on health in São Paulo. Information on their temporal coverage, spatial unit of analyses, exposure, lags, outcome and modifiers investigates, and study design are provided. They are also classified depending on the overall aim of the study as: investigating overall relationships (overall), extreme temperatures (extremes), vulnerable groups (vulnerability), UHI (UHI), using novel methodologies (novel methods) or exploring future projections (projections). In blue, colored if they part of the SALURBAL or MCC multi-city studies.

Reference	Location	Temporal coverage	Multi city or uni-city	Exposure	Lags	Outcome	Modifiers: description	Study design	Overall	Ex-tremes	Vulner-ability	UHI	Novel mde-thod	Projec-tions	SALUR-BAL	MCC
Sera2019	340 cities worldwide	1985-2014	Multi-location (unit: cities)	Daily mean temperature	Lag0-21	All non-external	Contextual [i.e., %population ¿65; life expectancy; GDP, labour productivity, educationa level, unemployment rate, gini index,m poverty gap, hospital-bed rates, type of surrounding region (rural/urban); urbanized area share (%), green area (m2 per million persons) and concetnration of population the the core (sprawl) + population density]	Time-series								
Gaspar-rini2017	451 cities globally (including São Paulo)	1984-2015 (city-specific)	Multi-location (unit: cities)	Daily mean temperature	Lag0-21	All non-external	Contextual [Country-level GDP]	Time-series								
Bakhtsiyar-ava2023	325 cities in South Ameirca	2002-2015	Multi-location (unit: cities)	Daily mean temperature	Lag0-21	All non-external and CVD	Contextual [i.e., Population (total and ¿65 years old), population density, built-up area, age-standardized mortality rate, living conditions score, secondary education, poverty, gini index and segregation]	Time-series								
Leiriao2023	Sao Paulo	1997-2007	Uni-city (unit: city)	Daily mean temperature,PM10, SO2, NO2, RH and wind speed	Lag0	CVD mortality among elderly	-	Artificial Neural Networks								
Bakhtsiyar-ava2022	265 cities in Brazil (including São Paulo), Mexico and Chile	2010-2015	Multi-location (unit: mu-nicipalities)	Daily mean temperature (population-weighted) assigned to 40 weeks prior to birth (exposure window)		Birthweight (in grams)	Invididual [i.e., Infant's sex, maternal education]	Time series								
Bell2008	Sao Paulo, Santiago de Chile, Mexico City	1998-2002	Multi-location (unit: cities)	Daily apparent temperature	lag0 and lag01	All non-external	Invididual [i.e., Age, sex, educational attainment]	Case-crossover								
Hajat2005	Sao Paulo, Delhi, London	1991-1994	Multi-location (unit: cities)	Daily temperature	Lag0 to lag 28	All non-external; cvd and resp	Invididual [i.e., Sex and age]	Time-series								
Sharovsky	Sao Paulo	1996-1998	Uni-city (unit: city)	Multiple exposures: weather variables (Daily mean temperature; relative humidity; atmospheric preasure), and air pollution (suffur dioxide, particulate matter, carbon monoxide)	Lag0-2 (moving average)	Myocardial infarction mortality	-	Time-series								
Gou-veia2003	São Paulo	1991-1994	Uni-city (unit: population groups)	Daily mean temperature	Lag0-1(heat); Lag0-20(cold); averaged	All non-external; cvd and resp	Invididual [i.e., Age as children, adults, eldery; sex] + Contextual [i.e., SES indicator as the arithmetic mean of: mean income level, average level of education, presence of adequate sewerage system in the household, percentage of households with piped treated water services, and crowding -district level]	Time series								
Bar-boza2022	Sao Paulo	2017	Uni-city (unit: census tracts)	Several enviornmental hazards including: air pollution, green space, and temperature.	-	All non-external	Contextual [i.e., IPVS 2010]	Health Impact Assessment								
Son2014	Sao Paulo	1996-2010	Uni-city (unit: city)	Daily mean temperature	Lag0-1 (heat) and Lag0-20 (cold)	All non-external, cvd and resp	-	Time series								
McMichael	Delhi, Monte Bucharest an	1989-1999 (country-dependent)	Multi-location (cities)	Daily mean temperature	Lag0-13 (cold) and Lag0-1 (heat)	All non-external mortality	Contextual [i.e., latitude, altitude, GNP, annual mean temperature, annual mean relative humidity, proportion of deaths in 65+ age group]	Time series								
Goncalves200	Sao Paulo	1996-2000	Uni-city (unit: city)	Daily mean temperature,PM10, SO2, CO, NO2, O3, relative and aboslute humidity, and thermal comfort index (or effective temperature).	Lag03 (moving average)	CVD mortality among elderly	-	Principal Component Analysis								
Sara2022	Sao Paulo	2006-2015	Uni-city (unit: city)	Extreme temperatures (12 heat wave and 9 cold spell definitions)	Lag0-10 (HW); Lag0-27 (coldspells)	All non-external causes, ischemic heart disease, cerebrovascular disease, ischemic stroke, hemorrhagic stroke and chronic obstructive pulmonary disease.	Individual [i.e., Sex, cause-specific death]	Time series								
Diniz2020	Metropoli-tan Region of São Paulo	2030-50 and 2079-99	Uni-city (unit: city)	Extreme temperature defined as period of at leas maximum temperatures above the threshold of t of the reference period (1996–2016)	Lag0-10	All cause, CVD and respiraotry among elderly	Invididual [i.e., Sex]	Time-series								
Arm-strong2003	Sao Paulo	1991-1994	Uni-city (unit: city)	Extreme temperature defined as number of degrees that the 2-day moving average was above 20degrees, zero if below 20degrees.	Lag0-2	All non-external causes	Contextual [i.e., SES]	Case-only approach								

2.8 Summary

Communities have evolved to adapt to their local climate through behavioral, technological and physiological changes. This has defined their optimal temperature range. Deviations from these optimal temperature lead to an increase in mortality and morbidity. Changes in the temperature brought by climate change, with more frequent and extreme temperature events, and in urbanicity with exposure to higher temperatures due to the UHI, are a real threat to population's health. A large body of evidence supports the non-linear and delayed association between daily mean temperature and morbidity and mortality. Yet, less is known about the spatiotemporal variation of the relationship at the local scale, in particular in urban settings, as well as of the role of individual and contextual determinants in modifying said relationship. In Chapter 3, I examined the temporal variability in the mortality risk to non-optimal temperatures using a flexible modelling framework to allow for non-linear and delayed effects. In Chapter 4, I predicted daily mean temperature at a high spatiotemporal resolution (daily, $500m^2$) using a RF algorithm. Finally, in Chapter 5, I investigated a broad myriad of contextual determinants of the temperature-mortality association using the temperature data generated in Chapter 4 and mortality records geo-coded at the residential address.

Chapter 3

Exploring disparities in adaptation to cold and heat

This chapter has been adapted from an open-access paper [216] which I first-authored. I hereby give credit to this work without citing the work in every instance in the text. The code used for the analysis can be accessed on my GitHub page [25] and an interactive version of the results is available on the project's website [26].

3.1 Overview and objectives

With exposure rapidly changing globally due to climate change, there is an interest in understanding whether populations are changing their susceptibility to temperature, in other words, whether they are adapting. In epidemiology, we understand adaptation as a reduction in the mortality risk associated with a given temperature profile. Two indicators that are often used to describe adaptation are: changes in the optimal temperature or MMT, which capture the comfort temperature of a population where the mortality is lowest, and changes in the risk associated to temperature above or below the MMT, expressed as the accumulated risk over all relevant lag days, also referred as cRR. In recent epidemiological studies, adaptation to higher temperature has been observed across several populations globally, with studies reporting a positive trend in MMT and a lowering in the cRR [11, 41, 84, 94, 95, 102]. Despite the mounting evidence, there is still a remarkable research gap when it comes to understanding the trends for cold-related mortality, specific vulnerable groups, in LMIC settings, and within urban settings [41, 95, 97]. Refer to Chapter 2 Section 2.2.6 for a detailed overview of the state-of-the-art literature on this.

Adaptation is believed to occur through changes in the level of exposure experienced by the population, and/or their vulnerability to its effects. Such changes are likely to vary over time, space and by population groups. Therefore, results from a certain area and time period are not necessarily

transferable or informative of the adaptive patterns of another time, location or population group. From a public health perspective, this has two consequences. On the one hand, it emphasizes the value of research exploring local contexts in order to provide relevant information to stakeholders and policy makers. On the other hand, it highlights the need for regular updates to ensure that the evidence is still representative of the target population, providing useful and adequate information for the development of suitable, timely and equitable prevention plans, such as HHWS and HHAP[32]. In developing HHWS, a common approach is to use a static temperature threshold above which, the associated mortality risk is considered to be excessive and preventable. According to the WHO guidelines [32], the gold standard for HHWS is to use epidemiological data to inform the thresholds of activation to be used. In order for HHWS to be effective, they have to capture the real risk of the population at a given time and set the thresholds accordingly. This thresholds are based on studies quantifying the cumulative mortality risk associated to temperature (cRR) in reference to the MMT. Thus, variations in the cRR and MMT not only are good indicators of adaptation, but can also flag the need for updates and improvements in public health measures such as HHWS. Finally, studies investigating historical adaptation patterns can lead to a better understanding of the underling drivers and mechanisms of adaptation which, in turn, can be used to inform future climate impact projections [96, 217].

In this chapter, I aim to provide a comprehensive analysis of the temporal trends in the temperature-mortality association in the State of São Paulo between 2000 and 2018, and to investigate the presence of disparities in the rate and magnitude of adaptation by sex, age, cause of death, ethnicity, and years of education. I use standard indicators for extreme and optimal temperature to provide an overview of adaptation, namely I explored both changes in the cRR and the MMT. In doing so, I provide information that can support policymakers and stakeholders to understand current temperature risks and their development in the last decades, identify vulnerable groups with poorer adaptation history, and inform interventions such as the development and implementation of HHWS with suitable thresholds to protect the population in São Paulo.

3.2 Data and Methods

3.2.1 Study area

The State of São Paulo is one of the 26 states that constitutes Brazil, located in the South-East. It expands over an area of 248,219.4km^2 and has approximately 39.9 million inhabitants as per the 2010 census (population density: 188.0/km^2) [23]. The Metropolitan Area of São Paulo (MASP) is located in the South-East of the state and it is composed of 39 municipalities, including the municipality of São Paulo (the state capital). It is home to 19,672,582 people which is about 30% of the state's population

(population density: 7,398.3 people. /km^2) [24]. According to the Köppen-Geiger classification, it has a humid subtropical climate in the North and a temperate oceanic climate in the South [193]. And so, São Paulo experiences a relatively dry season from June to August and a hot, rainy season from November to March.

3.2.2 Data sources

3.2.2.1 Mortality data

State-wide daily mortality counts with non-external factors as the primary cause of death were obtained from the national Public Health System database (*Sistema Único de Saúde, SUS*), run by the Ministry of Health of Brazil[218]. Data were checked for incorrect or missing age, sex or date fields (n=0), immortal years defined as ≥120 years old (n=0) and outliers, defined as any counts with a value above 2-fold the standard deviation (n=0). I investigated the following causes of death, as defined by the International Classification of Disease 10^{th} edition (ICD-10): (i) all non-external causes (ICD-10: A00-R99); (ii) CVD (ICD-10: Chapter I), and (ii) respiratory diseases (ICD-10: Chapter J). Death certificates in Brazil are compulsory. Since 1996, cause of death is coded according to the ICD-10, ensuring continuity in the coding system through my study period. Daily mortality counts were obtained stratified by age, gender, educational level and ethnic group providing further information on potential vulnerable groups.

3.2.2.2 Temperature data

Hourly temperature (°C) data were obtained from the Institute of *Instituto Astronomia, Geofísica e Ciencias Atmosféricas, Universidade de São Paulo* (IAG-USP) meteorological station[219]. The IAG-USP meteorological station is located in *Parque CienTec* at the main campus of the University of São Paulo, which falls within the *Parque Estadual dos Fontes do Ipiranga* in the *Agua Funda* neighborhood (coordinates: 23,6512°S, 46,6224°W; elevation: 799.2m; Figure 3.1). It collects information of a wide range of meteorological variables covering a period between 1997 to present. Air temperature is recorded hourly by an observer using an Assmann psychrometer and complemented with readings from a Daily Bimetallic Thermograph [220]. I checked all data for inconstancy and data quality. Finally, I aggregated hourly temperature to daily values using the 24h arithmetic mean. Alternative sources of temperature were explored, namely the ERA5-Land re-analysis gridded dataset. Upon exploration, the IAG-USP meteorological station stood out as the best choice for studying temperature-mortality association temporal trends in São Paulo. Later in this chapter (Section 3.2.4.3), I provide a detailed description of this investigation, including a sensitivity analyses comparing the mortality risk estimates from both temperature data sources.

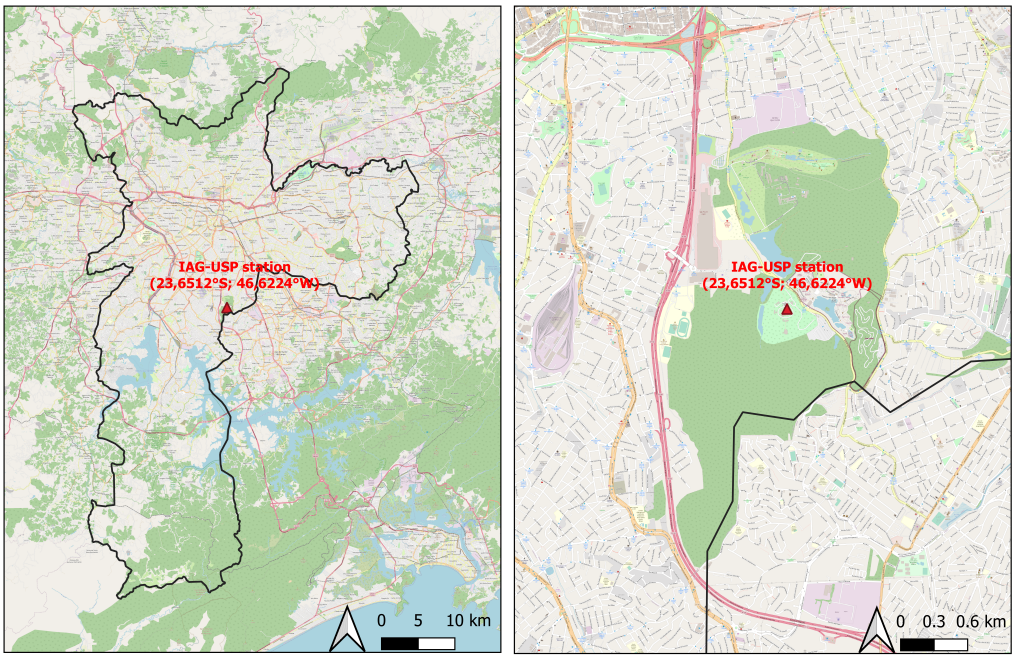


Figure 3.1: **Location of IAG-USP meteorological station.** (Left) Location of the IAG-USP meteorological station (red triangle) inside the municipality of São Paulo (black outline). (Right) Exact location of the station within the *Parque Estadual dos Fontes do Ipiranga*. Exact coordinates are provided in brackets.

3.2.2.3 Confounder data

The body's physiological response to extreme temperatures is closely related to evaporation processes which in turn, are linked to Relative Humidity (RH). RH in São Paulo shows little daily and seasonal fluctuation and thus, I would expect its impact to be minor. Nevertheless, to confirm that RH was not confounding the association between temperature and mortality, I adjusted for RH in my model, isolating this way the effects of temperature. Data on RH (%) was collected from the IAG-USP meteorological station (refer to 3.2.2.2 for more detail; Figure 3.1). Hourly RH was calculated from hourly wet bulb temperature and atmospheric pressure as observed in a barometer, complemented with estimates from a hygrograph with daily rotation [220]. Data were checked for inconstancy and quality. Hourly RH were aggregated to daily values using the 24h arithmetic mean.

Air pollution has also been suggested as a factor that may modify and/or interact with temperature [221]. On the one hand, air pollution dispersion patterns and concentrations are weather-driven. On the other hand, physio-pathological changes caused by air pollution could modulate the biological response to temperature, and vice versa [222]. Finally, analogous health outcomes and vulnerable populations have been independently identified for both exposures [223]. My focus here is to describe the effects of temperature independent of air pollution and so, I treated air pollution as a confounder. Air pollution in the State of São Paulo is monitored by the regional Environmental Agency (*Companhia Ambiental do Estado de São Paulo* (CETESB)) [224]. Currently, the network comprises 26

manual, 61 automatic and one mobile stations scattered across 35 municipalities[224]. The State of São Paulo is divided in 22 *Unidades de Gerenciamento de Recursos Hídricos* (UGRHI)[225] which are the working units for CETESB to manage their air monitoring network. For consistency with the spatial coverage of the meteorological data, I focused in the UGRHI located in the vicinity of the meteorological station. This consisted of 36 automatic stations (located in UGRHI number 6, yellow in Figure 3.2) which record Particulate Matter of 2.5micrometers ($PM_{2.5}$) and of Particulate Matter of 10micrometers (PM_{10}), sulfur dioxide (SO_2), nitrogen dioxide (NO_2), carbon monoxide (CO) and Ozone (O_3). $PM_{2.5}$ and PM_{10} capture a wide range of air pollution sources and so, they are often used as generic measures of air quality. The quality and availability of data for the $PM_{2.5}$ indicator was sub-optimal, and thus, I used PM_{10} as indicator of air quality in this study. Note that this will be the indicator used hereafter in this thesis when exploring air pollution. I downloaded the data for the stations that were either within or nearby the boundary of the State of São Paulo (n=35, Figure 3.2) as information could be borrowed from nearby areas if needed.

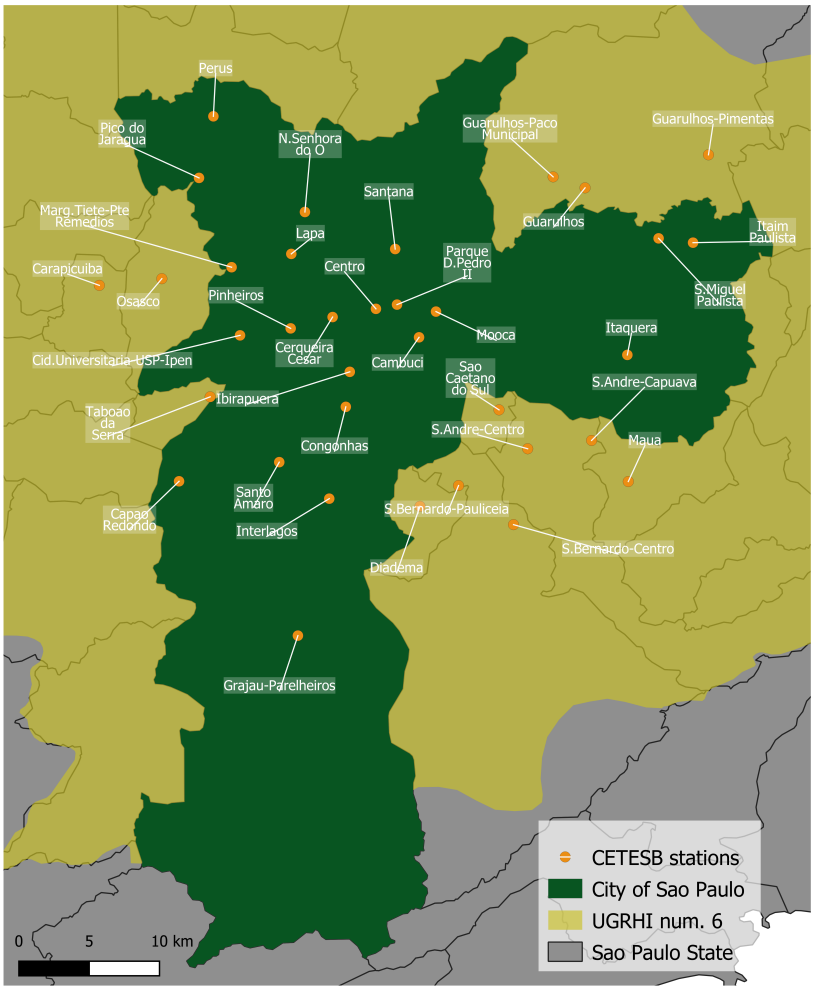


Figure 3.2: **Air pollution monitoring station network.** Map showing the location of the CETESB automatic stations (orange dots) selected for their proximity to the meteorological station. All stations belong to the UGRHI number 6 (yellow).

Out of the 35 stations, 32 collected PM_{10} measurements between 2000 and 2018. I performed a thorough data completeness check which consisted of the following two criteria: (i) at least 75% of valid hourly measurements, to ensure robust daily estimates, and (ii) at least 75% complete daily estimates across the study period to guarantee extensive and continuous temporal coverage. If the first condition was not met, the value was set to missing. If the second condition was not met, the station was discarded (n=11). At the end, I retained 18 stations for PM_{10}. For each station, I calculated the daily mean PM_{10} concentrations. Days with <75% valid hourly data were set to be missing.

Table 3.1: **Description of air pollution network spatial representativeness.** CETESB classification of air pollution monitoring stations based on their spatial representativeness

Representativeness	Description
Micro-scale	Representative for a radius of 100m.
Medium scale	Representative of urban areas of several blocks with similar characteristics (*quarteirões*), with a radius between 100 and 500m.
Neighbourhood scale	Representative of entire neighbourhoods within a city (*barrios*) containing relatively similar land usage, a radius between 500m and 4km.
Urban scale	Representative of an entire urban area, with a radius between 4 and 50 km. This scale normally requires more than one monitoring station to be defined.
Regional scale	Representative of a region, usually a rural area, of reasonably homogeneous geography and covering between 10s and 100s km.
National scale	Representative of an entire country.

These stations are often deployed to monitor specific local sources, e.g., industrial areas. This could result in biased estimates. In 2013, the Statutory Decree n° 59113, 23^{rd} April 2013article n°5 paragraph 4 [226], made mandatory to classify stations based on their spatial representativeness[227] (Table 3.1). For a given pollutant, the spatial representativeness is given by the amount of area around the monitoring station which shows a relatively similar concentration to that recorded by the station. This classification must consider: the type of pollutant; the proximity to pollutant sources both fixed and mobile; the intensity of the emissions; the topography; meteorological conditions of the area, and the altitude and position of the sampling. Urban scale stations are the best representation of entire urban areas, reflecting concentrations of areas between 4 and 50km, followed by neighbourhood scale stations covering between 500m and 4km. Thus, I used this as my reference stations and averaged the daily values across stations that had a pairwise Pearson correlation coefficient of $r \geq 0.90$ with any "urban scale" or "neighbourhood scale" reference station, and/or a $r \geq 0.85$ with two or more of these. In the case of PM_{10}, there was no "urban scale" station and five 'neighbourhood scale" stations. Based on the above criteria, I selected nine stations; five "neighbourhood scale"; three stations with $r \geq 0.85$ with more than two "neighbourhood scale" stations, and one with $r \geq 0.90$. Figure 3.3 shows the correlation matrix and selected stations. Finally, I averaged the 24h mean across all selected stations.

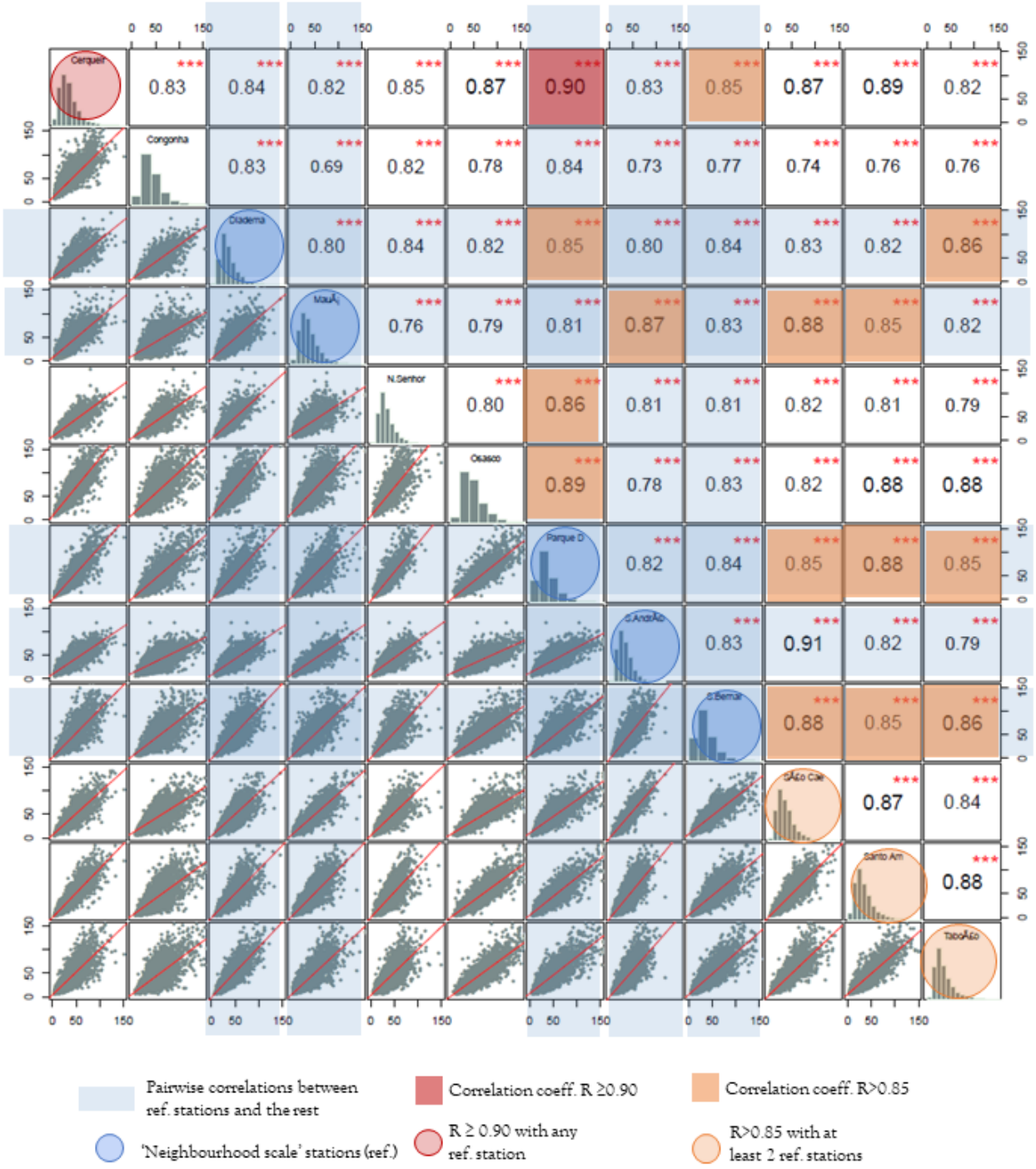


Figure 3.3: **Pairwise correlation of PM_{10} estimates across stations.**Pairwise correlation of PM_{10} estimates across all stations with valid data according to the completeness criteria. On the top quadrant, the correlation coefficients (***, significant); on the bottom quadrant the scatterplots and fitted linear regression (red line). Shaded in yellow, the pairwise correlations between the 'urban scale' reference (orange circles) stations and the rest; in red those with R>0.90, and circled in blue the stations selected for having R>0.90.

3.2.2.4 Vulnerability data

In order to investigate the presence of disparities in the rate and magnitude of adaptation, I collected information on potential individual determinants (effect modifiers), and quantified the temperature-mortality association independently for each category of individual determinant. For this study, I extracted information on age, sex, ethnic group, and years of education. When possible, I explored the joint effect of two determinants by creating sub-categories, for example, exploring age groups between sexes.

In Brazil, death certificates systematically collect the cause, date and location of death, as well as information on the age, sex, ethnic group, civil status and years of education of the deceased. The proportion of missing values of civil status and location of death was above 50%, and so, they were not considered in the analyses. For age, I focused on the two oldest groups (*65-79years old* and $\geq$ 80 *years old*) because of the higher vulnerability among the elderly reported in the literature [228]. Difference in the risk of cold and heat effects have also been reported by sex [13]. Here, I used a binary sex classification (*females* and *males*) as a more inclusive classification was not available. Information on ethnicity was provided as per the 2010 census coding: white (census term: *Branco*), Black (*Preto*), Mixed ethnicity (*Pardo*), East Asians (*Amarelo*) and Indigenous (*Indígena*). In order to ensure sufficient statistical power, I combined Black (*Preto*) and Mixed ethnicity (*Pardo*) group into a *non-white group*. East Asians (*Amarelo*) and Indigenous (*Indígena*) have been shown to have different risk profiles and so, were excluded from this group. Unfortunately, the counts were too small (67,782 cases combined) to be considered individually. Information on years of education was of poor quality (average: 35.6% missing values vs <5% for other covariates, Table 3.2). Hence, it was excluded from the main analyses. Upon investigation of the proportion of annual missing education values, I observed an improvement after 2011 (mean: 24.7%; $\min_{2018}$: 20.5; $\max_{2012}$:28.6%) compared to before 2011 (mean: 43.3%; range: 30.0-50.8%) (Table 3.3). So, I used data post-2011 to explore whether the temperature-mortality association varied by group of years of education (Table 3.3). I aggregated the total years of education into three categories to ensure contrast, i.e., $\leq$ 3 *years, 3 to 11 years*, and $\geq$ 11 *years*. Due to the smaller sample size, it was not possible to investigate the joint effect of years of education by categories of other individual determinants. Records with missing data were excluded from group-specific analyses after ensuring absence of bias through a chi-square Pearson statistic test.

Table 3.2: **Missing data per population group.** Percentage of records with missing and/or erroneous information, which were consequently excluded from each stratified analyses.

Data	Missing (%)
Gender	0.0%
Age group	0.3%
Ethnic group	3.7%
Education	35.6%
Gender and Age	0.3%
Gender and Ethnic group	3.7%
Age and Ethnic group	4.0%
Gender and Education	35.6%
Age and Education	35.6%

Table 3.3: **Missing data by year for variable "years of education".** Percentage of missing data on years of education by year, between 2000 and 2018.

Year	Missing (n)	Total deaths	Missing Percentage (%)
2000	121,315	238,959	50.8%
2001	117,243	235,987	49.7%
2002	112,544	237,741	47.3%
2003	112,134	240,253	46.7%
2004	111,112	243,984	45.5%
2005	106,879	236,456	45.2%
2006	106,633	243,955	43.7%
2007	102,264	244,653	41.8%
2008	103,382	249,247	41.5%
2009	103,556	56,627	40.4%
2010	97,771	264,951	36.9%
2011	80,802	270,367	29.9%
2012	77,335	270,432	28.6%
2013	77,092	276,980	27.8%
2014	76,542	281,624	27.2%
2015	71,369	287,645	24.8%
2016	67,992	296,359	22.9%
2017	63,867	294,753	21.7%
2018	61,034	298,313	20.5%

3.2.3 Statistical analyses

To assess the temporal trends in the temperature-mortality association in São Paulo between 2000 and 2018, I estimated the association for each year of the series through a quasi-Poisson regression and Time-Varying Distributed Lag Non-linear Models (tv-DLNM). Below, I describe the rational for choosing the tv-DLNM statistical framework, the model specification and the indicators of adaptation used in this study.

3.2.3.1 Trends in daily mean temperature

I visually investigated the temporal variability in temperature distribution, the trends in the annual 1^{st} and 99^{th} percentile of temperature, representing extreme cold and heat, and the Annual Mean Temperature (AMT). I formally tested for trends using linear regression with year as the independent variable. I also calculated summary statistics (mean, median, maximum, minimum, standard deviation (SD) and IQR) for the overall study period as well as for the early (2000-2004) and late years (2014-2018) to be able to see whether there were any differences.

In addition, I also investigated the potential impact of the El Niño Southern Oscillation (ENSO) phenomenon on the observed trends. In normal conditions, the so-called trade winds blow east to west carrying with them large amounts of warm water from the South American coast to Asia. During this process, cold water emerges from the depths, the so-called up-welling. ENSO is a natural phenomenon by which these normal conditions are altered. It is composed of two opposite events; El Niño and La Niña. During La Niña, trade winds are stronger than usual and push more warm water away from the South American coast, bringing in more cold water than usual. For the east Pacific coast, this results in areas of high pressure and reduced rainfall, sometimes causing extended droughts, and lower than average temperatures. During El Niño, the trade winds are weaker and so less warm water is pushed west. For areas in the east Pacific, this results in areas of low pressure and increased precipitation. The associated flooding increasing in severity if they follow an extended drought brought on by La Niña [229]. Despite these patterns can be felt worldwide, the most direct effects on regional climate are felt by the areas in the west coast of South America and the east coast of the Asian continent. Yet, when these events are severe, their effects can have a remarkable impact of further away places, such as São Paulo (east coast of South America). Thus, these patterns may influence the temperature-mortality trends and are worth investigating. Accordingly, I explored the concurrence of such events and assessed their correlation with the temperature trends recorded during the study period. Information on severe El Niño and La Niña events was collected from the NASA NOAA Climate Prediction Service of the National Weather Service Climate Platform [229].

3.2.3.2 tv-DLNM: overview of the modelling framework

I assessed the temporal trends in the annual temperature-mortality association through a tv-DLNM combined with a quasi-Poisson generalised linear regression model.

The standard DLNMs assume a time-constant exposure-lag-response association throughout the entire study period. In contrast, the tv-DLNM [76] approach, and extension of the standard DLNM, allows the temperature-mortality association to vary over time, making it suitable for this study. Statistically, it does so by adding a linear interaction term between time and the cb function of temperature to allow the exposure-lag-response association to vary over time in a linear manner. In statistics, an interaction occurs when an independent variable (X_1) has a different effect on the outcome (Y) depending on the value of another independent variable (X_2). Here, we are saying that the effects of temperature (X_1) on mortality (Y) vary depending on the value of time (X_2). For two continuous independent variables, interaction is expressed as the product between X_1 and X_2. In this case, this is the product between the function of temperature and time: $cb_{temp_t} \times time$. The results of the interaction models are interpreted conditionally to the values of X_2, in this case *time*. That is, the main coefficient for X_1 represents the effect of X_1 on Y for the value corresponding to the baseline value of X_2 used in the interaction term. Centering refers to the act of redefining the baseline or reference point used for this conditional interpretation. This results in exposure-lag-response associations specific for different points in time depending on the centering point selected.

In this study, I used a quasi-Poisson regression with a tv-DLNM formulation to capture non-linear and delayed associations changing over time. The model assumed the following general form:

$$log[E(Y_t)] = \alpha + cb_{tempt} + \sum_{k=1}^{7} \beta_k I(dow_t = k) + \gamma holidays_t + \delta PM10_{lag0-2_t} +$$

$$ns(RH_{lag0-2_t}, 3df) + ns(time, 10df \times year) + cb_{temp_t} \times time; \text{ for } t = 1...T = 6,940 \qquad \text{(Eq.1)}$$

where Y_t denotes the mortality counts at day t, α is the intercept and cb_{temp} is the bi-dimensional crossbasis matrix. The crossbasis function of daily mean temperature was composed by a natural cubic spline (ns) with one knot at the 75^{th} percentile of the temperature distribution for the exposure dimension, and a ns with an intercept and three knots equally spaced over the log scale of the temperature distribution for the lag dimension. I set the maximum lag effect to expand over 21 days to capture the delayed effects and short-term harvesting[53].

To eliminate potential confounding by time-varying factors, I included the following covariates: a time variable to account for long-term trends and seasonality modelled as a ns with $10df$ per year [$ns(time, 10df \times year)$]; categorical variables for day-of-the-week [dow], and national and state-wide holidays [$holidays$], and the *Natural Spline (ns)* of 2-day moving average of the daily mean RH with

$3df$ $[ns(RH_{lag0-2_t}, 3df)]$ and the 2-day moving average of the daily mean PM_{10} value $[PM_{10_{lag0-2_t}}]$.

Finally, to allow for time-varying exposure-lag-response associations, I added a linear interaction term between the temperature crossbasis variable and time$[cb_{temp_t} \times time]$. It represents the exposure-lag-response relationship at the centring point of the time variable used in the interaction term. Here, the interest is in the annual changes in the temperature-lag-mortality association; thus, I used the central day of each year (1^{st} July) to obtain equidistant estimates to summarize the annual variation. This is a common approach used in temperature-mortality studies[99, 230]. More information on the methodology can be found in [40]. To formally assess the significance of the interaction, I used a multivariate Wald test on the interaction coefficients, considering the associated (co)variance matrix and assuming a multivariate normal distribution[95]. The null hypothesis is that all coefficients are equal to zero, meaning that there is no change. In addition, I explored the visual representation of the interaction terms.

This model was the result of fitting competitive analyses with alternative model parametrization and selecting the best model fit based on the quasi-Akaike Information Criterion (qAIC)[231]. The qAIC is an estimator of prediction error often used to assess the relative quality of statistical models. I performed my analyses by cause of death (CVD and respiratory diseases) and for all non-external causes of death stratified by sex, age, ethnicity, and combinations of those covariates.

3.2.3.3 Indicators of adaptation

As discussed in Chapter 3 Section 2.2.3, the mortality risk associated to temperature follows a U-shaped distribution, with the lowest point or the temperature at which a given population shows the lowest risk of death, referred to as MMT. The slope of the curve defines the increment of risk per unit of temperature change, the steeper the slope the larger the increase in risk. Hence, a decrease in the slope (or flattening of the curve) over time indicates a reduction in the risk increment per unit of temperature change (Figure 3.4 (a)). The temperature-mortality risk curve may, therefore, be used as an indicator of the degree of adaptation to deviance from the MMT. The MMT itself usually shows a good correlation with the local AMT and summer mean temperature and so, it is considered a good indicator for long-term adaptation to the local usual temperatures. Therefore, shifts in the MMT over time have been previously used as an indicator of long-term adaptation to temperatures (Figure 3.4 (b)). These two events can happen individually or simultaneously.

Thanks to the flexibility of the DLNM, the effects of temperature can be explored in the continuum of the lag dimension. Here, I am interested in the total effects of temperature on mortality, that is, the cumulative effect over the relevant lags. Thus, here I compare the cumulative effect of temperature on mortality over 21 days, which I refer to as the cRR. The risk estimates are centered using as a reference the group and year-specific MMT. For each group and year, I plotted the cRR and compared

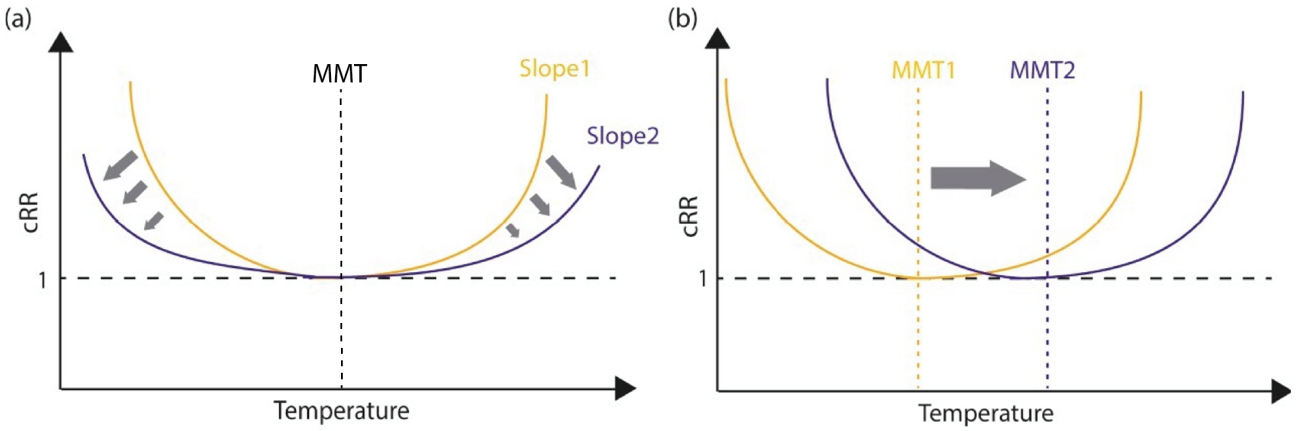


Figure 3.4: **Diagram of adaptation indicators.** Diagram of the two indicators of change used in this study. Given two imaginary time points (1, yellow and 2, purple), adaptation can be observed involving two changes: (a) A flattening in the cRR for non-optimal temperatures. In this case, it shows a reduction in risk for both heat and cold, to the left and right of the MMT (dashed vertical line), respectively. (b) A change in the period-specific MMT. In this case, it shows a displacement to the right, suggesting that the optimal temperature has increased.

them visually. In addition, I provide a numerical summary of the cRR (95% CI) for mild and extreme heat (90^{th} and 99^{th} percentiles, respectively) and cold (10^{th} and 1^{st}percentiles, respectively).

To calculate the MMTs, I considered only temperatures above the 1^{st} and below the 99^{th} percentile of the group-specific temperature distribution, known as *percentile constricted MMT*, to avoid unstable cRR curve tails affecting the computation of the MMT [9]. Hereafter, when I talk about MMT, I am referring to the percentile constricted MMT. In addition to the MMT, I also estimated the associated 95% empirical CI and the standard error using a parametric bootstrap estimator[9]. For each year, I calculated the specific MMT.

For the MMT as well as the cRR associated to mild and extreme heat and cold, I evaluated the temporal trends using a Weighted Least Square (WLS) regression with year as the independent variable. This method corrects for non-constant variance by weighting each observation by penalizing observations with large variance. I evaluated the goodness of fit through the Coefficient of determination or R-squared (R^2), calculated the Snedecor F-value and inspected the residuals versus the fitted values to assess the plausibility of the linearity assumption.

3.2.3.4 Investigating if the observed trends are associated to changes in climate

In order to understand adaptation, it is important to know whether the observed trends are associated to changes in the weather (i.e., climatic triggers) or independent of those (i.e., external triggers). Previous studies have shown that MMT is associated to the Annual Mean Temperature (AMT), indicating that long-term exposure to a similar weather, as captured by the AMT leads to changes in the comfort temperatures [9, 89]. Similarly, changes in exposure to extreme events, may lead to changes in the risk associated to those. Therefore, I investigated the association between the annual

MMT and AMT, cRR for extreme heat and annual 99^{th} percentile of temperature, and cRR for extreme cold and the 1^{st} percentile of temperature, using the same method. For each pair, I provide the regression coefficients, Pearson Correlation coefficients and p-value. Finally, I also investigated the possibility for strong El Niño and La Niña events to have influenced the results by comparing findings from ENSO years with non-ENSO years.

3.2.4 Sensitivity analyses

To evaluate the impact of model choices in my results and the robustness of the findings, I performed several sensitivity analyses. These included primarily changes in the model parametrization, modification in the specification of the RH and PM_{10}, and the use of gridded datasets for the temperature assessment instead of a reference monitoring station. For each of these, I calculated the model performance using the qAIC score and estimated the resulting MMT and cRR for all non-external causes of death with and without interaction. These results were then compared to the main model. Below, I describe each of them and the motivation behind.

3.2.4.1 Model parametrization

I performed a total of 23 sensitivity analyses which assessed the sensitivity of the model to changes in the parameter specifications. These included changes in: the lag dimension [maximum lag values: 15, 21 and 24, and number of knots: 2 and 3]; the temperature-mortality dimension [the *df*: 2 and 3, and location of the knots: 75^{th}; $50^{th}$and 90^{th}; 25^{th}, $50^{th}$and 90^{th}; 25^{th}, 50^{th}, $90^{th}$and 99^{th}; 10^{th}, $75^{th}$and 90^{th}], the seasonality [number of *df* per year: 3, 7, 8, 9, 10]. Refer to Table 3.4 for details.

3.2.4.2 Relative humidity and air pollution

To test the impact of using different parameterization of the RH variable, I ran the analyses adding RH on the same day (RH_{lag0}), moving average over lag0-1 (RH_{lag0-1}) and between lag0-2 (RH_{lag0-2}), all adjusted as a natural cubic spline with 3 *df* , and RH_{lag0} adjusted as a natural cubic spline with 4 *df*. The range of parameters used was informed by the literature. Similarly, I tested the model performance using different parametrization of PM_{10} daily concentration variable, informed by the literature, namely using the moving average over the day of the event and the previous day ($PM_{10lag0-1}$), two days ($PM_{10lag0-2}$) and three before ($PM_{10lag0-3}$). The details on the model parametrization can be found in Table 3.4.

Table 3.4: **Sensitivity analyses.** Sensitivity analyses computed on all-non external causes only by varying modelling choices, and controlling for PM_{10} and relative humidity. qAIC values are provided for each sensitivity analysis. Model of choice in bold

SA	Exposure basis	Exposure basis knots	Lag basis	Lag basis knots	Lag max	*df* seasonality	RH term	PM_{10} term	$qAIC_{noint}$	$qAIC_{int}$
1	Quadratic B-spline	1 knot at 75^{th} percentile	cubic B-spline	3 knots equally spaced log scale + intercept	21	ns(time, *df*=10/year)	-	-	66709.4	66468.7
2	Cubic B-spline	1 knot at 75^{th} percentile	cubic B-spline	3 knots equally spaced log scale + intercept	21	ns(time, *df*=10/year)	-	-	66684.6	66475.3
3	Cubic B-spline	1 knot at 90^{th} percentile	cubic B-spline	3 knots equally spaced log scale + intercept	21	ns(time, *df*=10/year)	-	-	66711.5	66475.4
4	Cubic B-spline	2 knot at 50^{th} and 90^{th} percentile	cubic B-spline	3 knots equally spaced log scale + intercept	21	ns(time, *df*=10/year)	-	-	66690.2	66495.8
5	Cubic B-spline	3 knot at 25^{th}, 50^{th}, 90^{th} percentile	cubic B-spline	3 knots equally spaced log scale + intercept	21	ns(time, *df*=10/year)	-	-	66699.8	66578.1
6	Cubic B-spline	4 knot at 25^{th}, 50^{th}, 90^{th}, 95^{th} percentile	cubic B-spline	3 knots equally spaced log scale + intercept	21	ns(time, *df*=10/year)	-	-	66684.6	66607.9
7	Cubic B-spline	3 knot at 10^{th}, 75^{th}, 90^{th} percentile	cubic B-spline	3 knots equally spaced log scale + intercept	21	ns(time, *df*=10/year)	-	-	66705.3	66614.9
8	Cubic B-spline	1 knot at 75^{th} percentile	cubic B-spline	2 knots equally spaced log scale + intercept	21	ns(time, *df*=10/year)	-	-	66823.5	66657.5
9	Cubic B-spline	1 knot at 75^{th} percentile	cubic B-spline	3 knots equally spaced log scale + intercept	21	ns(time, *df*=10/year)	-	-	66684.6	66657.5
10	Cubic B-spline	1 knot at 75^{th} percentile	cubic B-spline	3 knots equally spaced log scale + intercept	21	ns(time, *df*=10/year)	-	-	66684.6	66657.5
11	Cubic B-spline	1 knot at 75^{th} percentile	cubic B-spline	3 knots equally spaced log scale + intercept	15	ns(time, *df*=10/year)	-	-	66882.1	66657.5
12	Cubic B-spline	1 knot at 75^{th} percentile	cubic B-spline	3 knots equally spaced log scale + intercept	24	ns(time, *df*=10/year)	-	-	66640.5	66657.5
13	Cubic B-spline	1 knot at 75^{th} percentile	cubic B-spline	3 knots equally spaced log scale + intercept	21	ns(time, *df*=10/year)	-	-	66684.6	66661.1
14	Cubic B-spline	1 knot at 75^{th} percentile	cubic B-spline	3 knots equally spaced log scale + intercept	21	ns(time, *df*=9/year)	-	-	66843.0	66661.7

qAIC, quasi-Akaike Information Criteria; ns, natural cubic spline; *df*, degrees of freedom; PM_{10}, particulate matter ¡10μm.

Table 3.4: (Cont.) Sensitivity analyses computed on all-non external causes only by varying modelling choices, and controlling for PM_{10} and relative humidity. qAIC values are provided for each sensitivity analysis. Model of choice in bold

SA	Exposure basis	Exposure basis knots	Lag basis	Lag basis knots	Lag max	*df* seasonality	RH term	PM_{10} term	$qAIC_{noint}$	$qAIC_{int}$
15	Cubic B-spline	1 knot at 75^{th} percentile	cubic B-spline	3 knots equally spaced log scale + intercept	21	ns(time, *df*=8/year)	-	-	66908.7	66669.9
16	Cubic B-spline	1 knot at 75^{th} percentile	cubic B-spline	3 knots equally spaced log scale + intercept	21	ns(time, *df*=7/year)	-	-	67148.2	66685.1
17	Cubic B-spline	1 knot at 75^{th} percentile	cubic B-spline	3 knots equally spaced log scale + intercept	21	ns(time, *df*=3/year)	-	-	68386.2	66685.6
18	Cubic B-spline	1 knot at 75^{th} percentile	cubic B-spline	3 knots equally spaced log scale + intercept	21	ns(time, *df*=10/year)	ns(RH_0, *df*=3)	-	66688.3	66689.9
19	Cubic B-spline	1 knot at 75^{th} percentile	cubic B-spline	3 knots equally spaced log scale + intercept	21	ns(time, *df*=10/year)	ns(RH_{lag01}, *df*=3)	-	66640.3	66707.8
20	Cubic B-spline	1 knot at 75^{th} percentile	cubic B-spline	3 knots equally spaced log scale + intercept	21	ns(time, *df*=10/year)	ns(RH_{lag02}, *df*=3)	-	66603.8	66798.0
21	Cubic B-spline	1 knot at 75^{th} percentile	cubic B-spline	3 knots equally spaced log scale + intercept	21	ns(time, *df*=10/year)	ns(RH_0, *df*=4)	-	66688.9	66820.9
22	Cubic B-spline	1 knot at 75^{th} percentile	cubic B-spline	3 knots equally spaced log scale + intercept	21	ns(time, *df*=10/year)	-	PM_{10_lag01}	66509.4	66863.1
23	Cubic B-spline	1 knot at 75^{th} percentile	cubic B-spline	3 knots equally spaced log scale + intercept	21	ns(time, *df*=10/year)	-	PM_{10_lag02}	66509.4	66890.6
24	Cubic B-spline	1 knot at 75^{th} percentile	cubic B-spline	3 knots equally spaced log scale + intercept	21	ns(time, *df*=10/year)	-	PM_{10_lag03}	66527.7	67119.0
Main model	Cubic B-spline	1 knot at 75^{th} percentile	cubic B-spline	3 knots equally spaced log scale + intercept	21	ns(time, *df*=10/year)	ns(RH_{lag02}, *df*=3)	PM_{10_lag02}	66501.7	68362.2

qAIC, quasi-Akaike Information Criteria; ns, natural cubic spline; *df*, degrees of freedom; PM_{10}, particulate matter ¡10μm.

3.2.4.3 Temperature gridded datasets

Re-analysis data products, such as the ERA5-Land data [117] developed by the ECMWF [115], have been recently proposed as a good source of temperature data for epidemiological studies [36]. Here, I investigated whether the choice of temperature data source could alter the findings reported here and discuss which temperature data source is more adequate in this situation.

For the purpose of this comparison, I used ERA5-Land data. The combines past short-range weather forecasts with *in situ* observations from weather stations through data assimilation, at a 9km^2 spatial resolution. It provides global coverage from 1959 to present. I aggregated the gridded data using population-weighting. This method takes into account the spatial distribution of the population over the area and so, should be a better proxy for population exposure than geometric centroids. This method assigns a higher weight to grid cells covering areas that are more populous and a lower weight to those that are less populous. For the population estimates, I used the 2010 census data. The census tract centroids were intersected with the ERA5-Land grid cells (9km^2) and the total population count per grid cell extracted. The grid specific population-weight was defined as the grid-specific population count divided by the total population in the state. See Eq.2 for the algebraic expression, where n denotes a grid cell out of the total ERA5-Land grid cells (m) intersecting the State of São Paulo. The temperature at each grid cell n for each day i denoted as $Temp_{n,i}$,was multiplied by its population weight, denoted as $Pop\ weight_n$. This was then added up to obtain a unique daily mean temperature estimate for the entire State of São Paulo, denoted as $Temp_i$ (Eq. Eq.3).

$$Pop \text{ weight}_n = \frac{Pop\ count_n}{\sum_{n=1}^{m} Pop\ count_n} \quad \text{(Eq.2)}$$

$$Temp_i = \sum (Temp_{n,i} \times Pop\ weight_n) \quad \text{(Eq.3)}$$

I compared the daily mean temperature from IAG-USP against the population-weighted daily temperature from ERA5-Land through visually assessing the agreement between their time series and density plots. To formally quantify their degree of agreement, I calculated the Coefficient of determination or R-squared (R^2) and the Root Mean Square Error (RMSE). Finally, I fitted the main model using the ERA5-Land population-weighted daily mean temperature estimates for the whole study period (overall, no interaction) for all non-external causes of death, as a sensitivity analyses.

3.2.5 Software

All analyses were performed in R software (version 3.6.3; R Development Core Team), mainly using the *dlnm* R package [77]. P-values <0.05 were considered statistically significant.

3.2.6 Data access disclaimer

Computer code for replicating the results can be found in my GitHub repository[25] . Interactive visualizations of the results are available on the project's website [26]. State-wide daily mortality records are publicly available for download at the national Public Health System database DATASUS platform [218]. Hourly weather measurements can be obtained upon request from the Institute of Astronomy Geophysics and Atmospheric Sciences and University of São Paulo (IAG-USP). Information on data access in their website [219]. Air pollution data can be accessed upon request from the CETESB website [224].

3.3 Results

3.3.1 Descriptive statistics

3.3.1.1 Temporal trends in meteorological and air pollution data

Table 3.5 provides summary statistics for daily mean temperature (°C), RH (%), and air pollution (PM_{10}, μ/m^3). Information is provided for all study period, as well as for the five first (2000 to 2014) and last years (2014 to 2018) of the study period to give an overview of changes in time. These tables are accompanied by Figure 3.5 which depicts the daily (Panel A) and annual (Panel B) oscillations in temperature, RH and air pollution.

Table 3.5: **Summary statistics meteorological and air pollution data.** Summary statistics for daily temperature (°C), air pollution (PM_{10}, $\mu g/m^3$) and RH (%) in São Paulo for all (2000-18), early (2000-04) and late (2014- 18) years of the study period.

	All period (2000-2018)	**Early (2000-2004)**	**Late (2014-2018)**
Temperature (°C)			
Mean	19.6	19.6	19.9
Median	19.8	19.9	20.0
Maximum	28.0	27.6	28.0
Minimum	7.3	7.9	8.8
SD	3.3	3.4	3.4
IQR	4.8	4.7	4.9
Air Pollution (PM_{10}, μ/g^3)			
Mean	36.2	43.5	29.8
Median	32.2	38.9	26.8
Maximum	156.2	156.2	95.6
Minimum	6.2	6.7	6.2
SD	18.1	21.0	13.6
IQR	21.9	26.3	17.3
Relative Humidity (%)			
Mean	80.0	79.6	79.7
Median	80.8	80.4	80.6
Maximum	97.3	97.3	96.3
Minimum	34.3	45.5	41.2
SD	8.5	8.5	8.7
IQR	10.5	10.5	10.7

RH = relative humidity; SD = Standard deviations; IQR = Inter-quartile range.

The average daily mean temperature between 2000 and 2018 was 19.6°C (min: 7.3°C, max: 28.0°C) (Table 3.5). The annual average daily mean temperature remained mostly stable over time, with fluctuations of ±2°C compared to the study period mean. The 1^{st} and 99^{th} percentile of daily temperature oscillated between years and showed a slight increase over time, although this trend was not statistically significant (data not shown). Finally, the daily mean temperature showed a clear seasonality (p-value<0.05), with the hottest months being December-March, and the coldest May-August (Figure 3.5).

As expected for its *humid subtropical* (North) and *temperate oceanic* climate (South), the RH in São Paulo was high with an annual average of 80.0% (SD: 8.5) (Table 3.5). Over the study period, the mean and 99^{th} percentiles of the annual distribution remained rather stable with no remarkable trend. This can be visually observed from Figure 3.5 Panel B and comparing the values in Table 3.5 for the early and late period, were we see barely any change. The extreme low annual values (i.e., 1^{st} percentile) were more volatile, yet no clear trend was visible. The daily recordings showed a very attenuated seasonality, which followed the dry (June to August) and rainy season (November to March), yet it was not statistically significant (p-value>0.05).

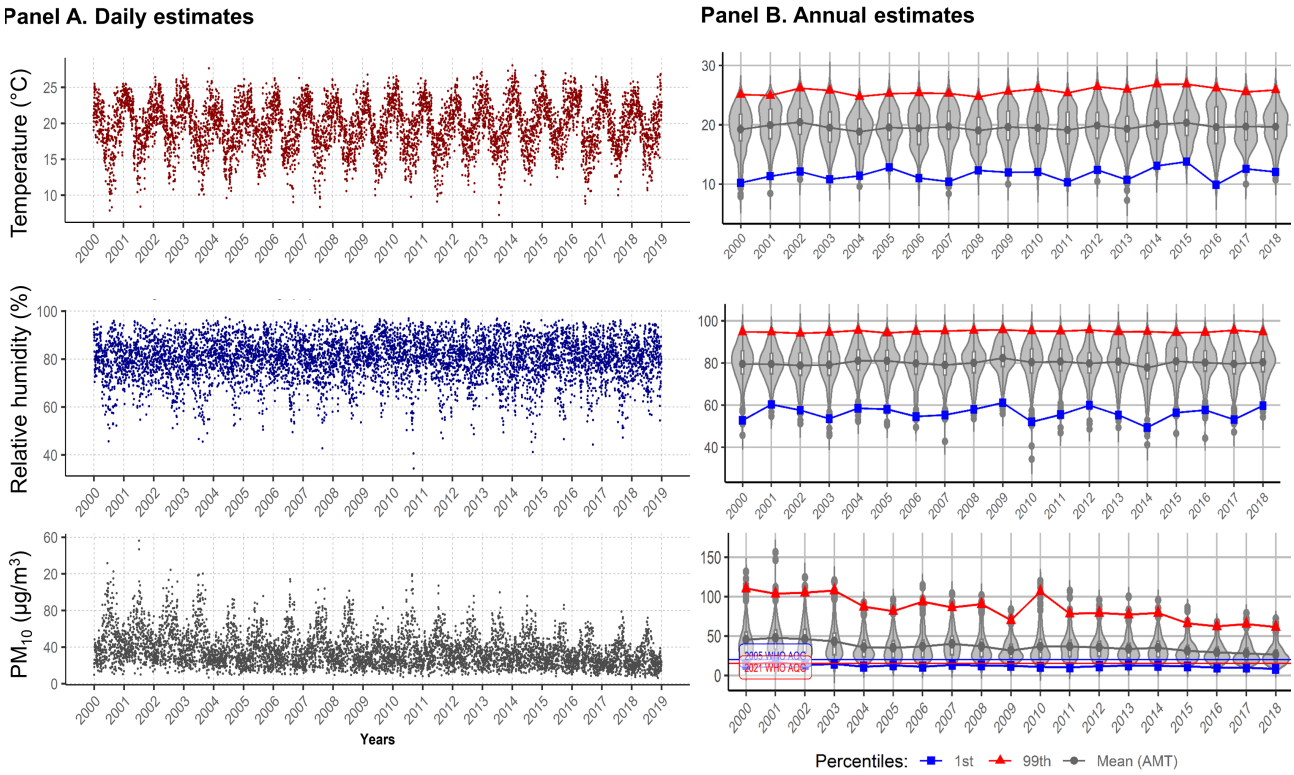


Figure 3.5: **Meteorological and air pollution daily and annual estimates.** Daily (Panel A) and annual average (Panel B) mean temperature (°C, top), relative humidity (RH,%, center) and PM_{10} daily mean concentrations ($\mu g/m^3$, bottom). Overlaid to Panel B, the 99^{th} percentile (red), the mean (grey) and 1^{st} percentile (blue). The 2005 and 2021 WHO Air Quality Guidelines for annual PM_{10} concentrations are shown as a blue and red solid line, respectively (Panel B, bottom).

Finally, the average daily mean PM_{10} concentration for the studied period was 36.2μg/m^3 (min: 6.3μg/m^3, max: 156.2 μg/m^3). Although annual average air pollution concentrations decreased over time from 45.09μg/m^3 in 2000 to 26.23μg/m^3 in 2018, levels continued to exceeded the 20μg/m^3 recommended by the 2005 WHO Air Quality Guidelines [232], enforced during the study period (Figure 3.5, Panel B bottom). Similarly, between 2000 and 2018, the 24h PM_{10} limits (50μg/m^3) were exceeded during a total of 19.14% days. Finally, the daily estimates fluctuated within a year, with higher values in the rainy season (November-March, Figure 3.5 Panel A).

According to the Climate Prediction Service of the National Weather Service [229], during my study period there were four El Niño events (2002-03,2005-06, 2009-10 and 2015-16) and three La Niña events (2007-08, 2010-11, and 2018-19) of relatively high strength (Figure 3.6). Despite this, I found no strong difference in the temperatures recorded by the IAG-USP meteorological station (Figure 3.6).

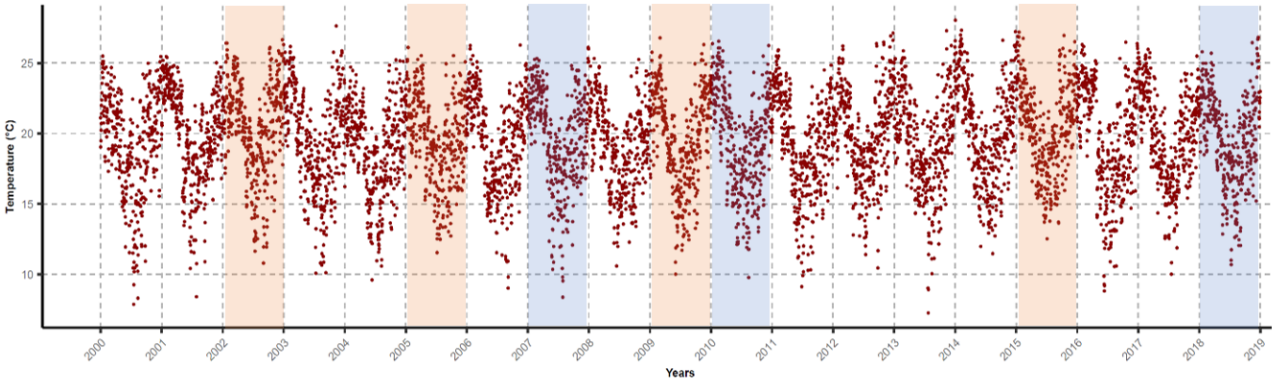


Figure 3.6: **ENSO-patterns and IAG-USP temperature data.** Time series of daily mean temperature as recorded by the IAG-USP meteorological station (red dots). Shaded in red the El Niño years, and in blue the La Niña years.

3.3.1.2 Temporal trends in mortality data

Figure 3.7 displays the variability in daily death counts for the same groups. Table 3.6 provides summary statics daily death counts for each population group across all study period for the five first and last years of the study period to inform about change over time. Overall, 4,471,000 deaths were recorded for all non-external causes of death, of which 1,492,164 were from CVD and 614,959 from respiratory diseases (33.4% and 13.8%, respectively). Of all non-external causes of death, 53.6% occurred among men, 70.0% among white population, 28.0% among the $\geq$ 80 years old and 14.9% among those with less than 3 years of education.

When comparing the first to the last five years, we see that the gender gap decreased over time so did the differences between the two age groups investigated. This can also be seen from the plots showing daily mortality counts increased (Figure 3.7). Looking at absolute mortality counts, we see that there is a positive trend across all groups, as expected considering the increase in population during the study period. By education, only people with 4 to 11 years and above 11 years of education experienced an increase in death counts, probably associated to an increase in literacy.

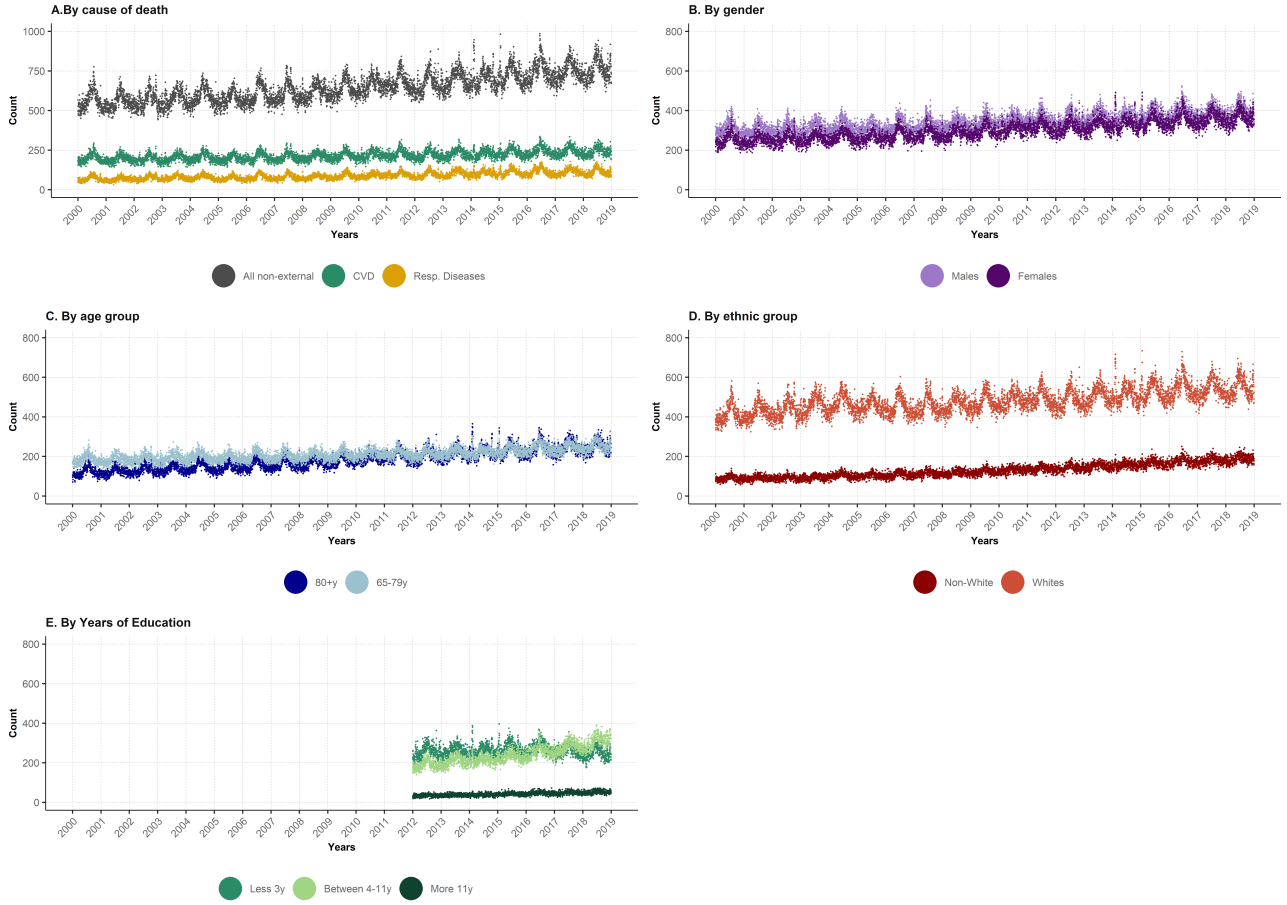


Figure 3.7: **Time series of death counts.** Daily death counts by: (A) cause of death, (B) gender, (C) age group, (D) ethnic group, and (E) years of education (only post-2011, see methods section for justification) recorded in São Paulo, Brazil (2000-2018). Note the y-axes are scaled to each population group's counts.

Table 3.6: **Summary statistics mortality data.** Mortality counts and percentage (%) by cause of death, age group, gender, and ethnic group in São Paulo for all (2000-18), early (2000-04) and late (2014- 18) years of the study period.

	All period (2000-2018)	**Early (2000-2004)**	**Late (2014-2018)**
All non-external	4,471,000	1,035,174	1,343,444
By cause of death$^{\gamma}$			
Cardiovascular	1,492,164 (33.4%)	361,195 (34.9%)	431,911 (32.1%)
Respiratory diseases	614,959 (13.8%)	132,594 (12.8%)	199,971 (14.9%)
By gender			
Females	2,074,790 (46.4%)	466,416 (45.1%)	639,714 (47.6%)
Males	2,395,487 (53.6%)	568,684 (54.9%)	703,471 (52.4%)
By age group$^{\gamma}$			
Aged 65-79y o	1,463,498 (32.7%)	346,241 (33.4%)	435,915 (32.4%)
Aged 80+yo	1,263,237 (28.3%)	240,325 (23.2%)	432,443 (32.2%)
By ethnic group$^{\gamma}$			
Non-Whites	896,669 (20.1%)	170,592 (16.5%)	315,624 (23.5%)
Whites	3,332,295 (74.5%)	800,232 (77.3%)	976,588 (72.7%)
By years of education$^{\gamma,*}$			
Less than 3y	668,220 (14.9%)	-	479,869 (46.4%)
Between 4-11y	614,751 (13.7%)	-	470,037 (45.4%)
More than 11y	109,124 (2.4%)	-	83,213 (8.1%)

$^{\gamma}$ Percentages estimated over the total population (including other groups not shown here and missing values) hence the percentages do not add up to 100%.

* Data shown only for years after 2011 due to high proportion of missing values (>30%) in prior years.

3.3.2 Disparities in adaptation

In this section, I discuss the main findings trends observed in the two adaptation indicators explored, namely the MMT and the cRR. For an overview of the temperature-mortality association for each population group across all study period (no interaction), please refer to Appendix Figure 6.2.

3.3.2.1 Minimum mortality temperature (MMT)

Table 3.7 contains the MMT with the 95%CI. The first thing to highlight, is the diversity in the temporal changes in MMT, both in terms of magnitude and direction (shifts to the right or left), observed across population groups. In general, the MMT tended to increase over time across all groups, with the exception of males aged 65-79 years old(-0.62); and non-white (-0.83), non-white population aged 65-79years old (-1.73) and females non-white (-5.26) (Table 3.7). The largest increase was among females aged ≥80 years old (+3.29); the lowest among among females non-white (-5.26). The MMT trends can be found in Figure 3.8 with the corresponding WLS regression equation, correlation coefficient and p-value, by population group. For all groups, I observed a statistically significant positive linear trend in the MMT estimates, except males aged 65-79 years old (WLS regression coefficient: -0.0323), non-white males (-0.0484), and non-white population aged 65-79 years old (-0.931) groups, which showed a decreasing trend. The largest increase was observed among females aged ≥80 years old (+0.210), and white females (+0.169). All trends were statistically significant

($p<0.05$, Figure 3.8). The MMT for each year can also be seen in Figure 3.9 as year-specific vertical lines for each population group.

Table 3.7: **MMT for 2000 and 2018 by population group.** Minimum mortality temperature (MMT) estimates for the first (2000) and last (2018) year of the study period and absolute difference between them (Δ MMT).

Population group		Year	MMT	95% CI	ΔMMT
Overall		2000	21.08	(20.52;28.05)	1.59
		2018	22.67	(21.72; 23.09)	
Gender	Females	2000	20.36	(19.78;28.05)	2.53
		2018	22.88	(21.93;23.19)	
	Males	2000	21.83	(20.98;23.09)	0.31
		2018	22.15	(21.08;23.09)	
Age group	65-75 years old	2000	22.03	(20.36;23.30)	0.20
		2018	22.23	(21.08;23.19)	
	≥80 years old	2000	20.52	(19.98;28.05)	2.36
		2018	22.88	(22.2;23.30)	
Ethnic group	White	2000	21.08	(20.61;28.05)	1.71
		2018	22.79	(21.72;23.19)	
	Non-white	2000	21.54	(20.26;28.05)	0.61
		2018	22.15	(7.25;28.05)	
Gender and Age group	Females aged 65-79y	2000	21.26	(20.08;28.05)	1.30
		2018	22.56	(7.25;23.43)	
	Females aged ≥80y	2000	19.68	(18.99;28.05)	3.29
		2018	22.98	(22.23;23.30)	
	Males aged 65-79y	2000	22.46	(7.25;23.30)	-0.62
		2018	21.83	(20.98;28.05)	
	Males aged ≥80y	2000	21.63	(7.25;28.05)	1.16
		2018	22.79	(20.52;23.43)	
Gender and Ethnic group	Females non-white	2000	28.05	(18.79;28.05)	-5.26
		2018	22.79	(7.25;23.67)	
	Females white	2000	20.36	(19.78;28.05)	2.62
		2018	22.98	(21.93;23.30)	
	Males non-white	2000	22.46	(20.16;28.05)	-0.83
		2018	21.63	(20.89;28.05)	
	Males white	2000	21.72	(7.25;28.05)	0.74
		2018	22.46	(20.89;23.19)	
Ethnic and Age group	Non-white aged 65-79y	2000	23.57	(21.17;25.42)	-1.73
		2018	21.83	(7.25;28.05)	
	Non-white aged ≥80y	2000	20.36	(17.55;28.05)	1.79
		2018	22.15	(20.26;23.57)	
	White aged 65-79y	2000	21.63	(7.25;28.05)	0.83
		2018	22.46	(20.89;23.30)	
	White aged ≥80y	2000	20.52	(7.25;28.05)	2.57
		2018	23.09	(22.56;23.30)	

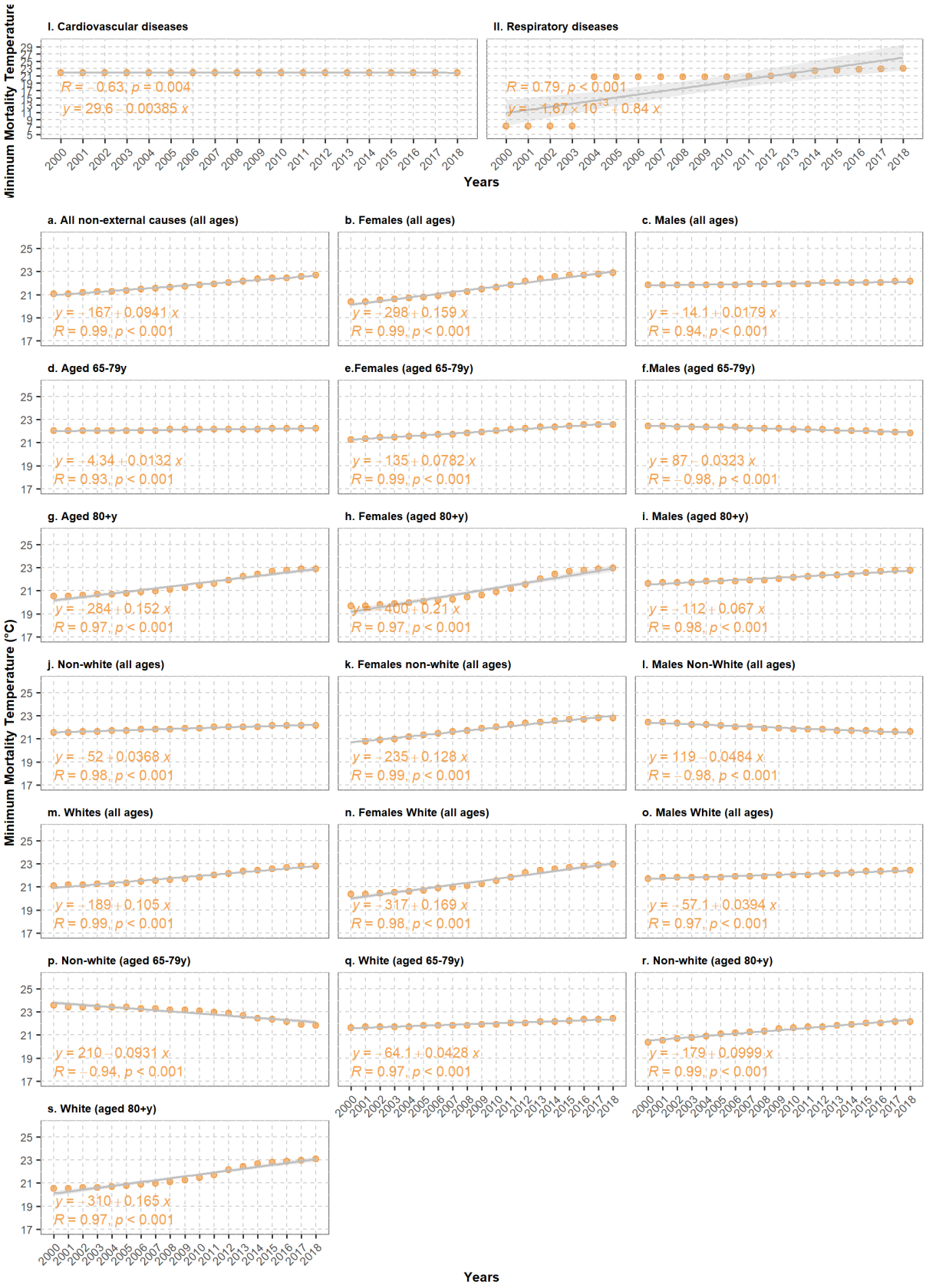


Figure 3.8: **Annual MMT estimates by population group.** Annual MMT estimates for (I) cardiovascular diseases (CVD) mortality and (II) respiratory diseases mortality, and for (A-S) all non-external causes of deaths by age, sex, ethnic groups, and combination of those. In orange, the fitted linear regression line, WLS regression coefficients, Pearson correlation estimate and p-values.

3.3.2.2 Cumulative relative risk (cRR)

Table 3.8 the cRR (95%CI) estimates at the 1^{st} (extreme cold), 10^{th} (moderate cold), 90^{th} (moderate heat) and 99^{th} percentiles (extreme heat) predicted for the first (2000) and last (2018) years. Formal assessment of the statistical significance of the temporal changes is provided by the Wald test of the interaction (marked with a * those that are statistically significant). Figure 3.9 displays the annual cRR curves with their MMTs (dashed vertical lines) by population group (models with interaction). Figure 3.10 shows the annual cRR associated to the 1^{st} (extreme cold) and 99^{th} percentiles (extreme heat) per population group with their 95% CIs. Figure 3.11 provides a visual representation of the interaction terms by population group. Finally, see Appendix Figure 6.2 for an overview of the temperature-mortality association across for each population group across all study period (no interaction).

As expected, the association follows a U-shape with risk increasing for extreme heat (red) and cold (blue) temperatures. Overall, I observed a statistically significant increase over time in cRR related to heat for all non-external and CVD-related deaths (Figure 3.10 top). The cRR associated with cold temperatures decreased for CVD deaths. For respiratory diseases deaths, the cRR profile changed remarkably, showing a larger risk to extreme cold, yet a reduced risk to mild heat and cold. For non-external causes of death by population group, the most remarkable finding was the heterogeneity in the trends (Table 3.8; Figure 3.10). The trends were different for heat than for cold. For heat (right side of the MMT in Figure 3.9),the eldest groups of population and females seem to experience an increase in the risk; whereas males and non-white population seem to experience a decrease. For cold temperatures (left of the MMT in Figure 3.9), all groups seem to experience a slight increase in the risk except for the non-white population subgroups who experience a slight decrease and females whose risk experience minimal changes. In general, all female sub-groups experienced an increase in the cRR for heat and a slight decrease for cold. The reverse was observed for male sub-groups, except for non-white males. The mortality risk to heat and cold increased among those aged $\geq$80 years old, whereas for those aged 65-79 years old, it only increased slightly for extreme cold. Finally, the non-white population showed a decreasing death risk associated with cold, which was not found among the white population. Overall, the time interaction term was statistically significant for all groups (Wald test p-value $<$0.05; Table 3.8), except for people aged 65-79 years old and non-white population, as well as for most of the subgroups. This was most likely due to the larger instability of the estimates due to smaller counts, which also reflects on larger 95% CI seen in Figure 3.11. Analyses by education level showed small and insignificant changes, with large and overlapping 95% CI (Appendix Figure 6.3) and were not explored further.

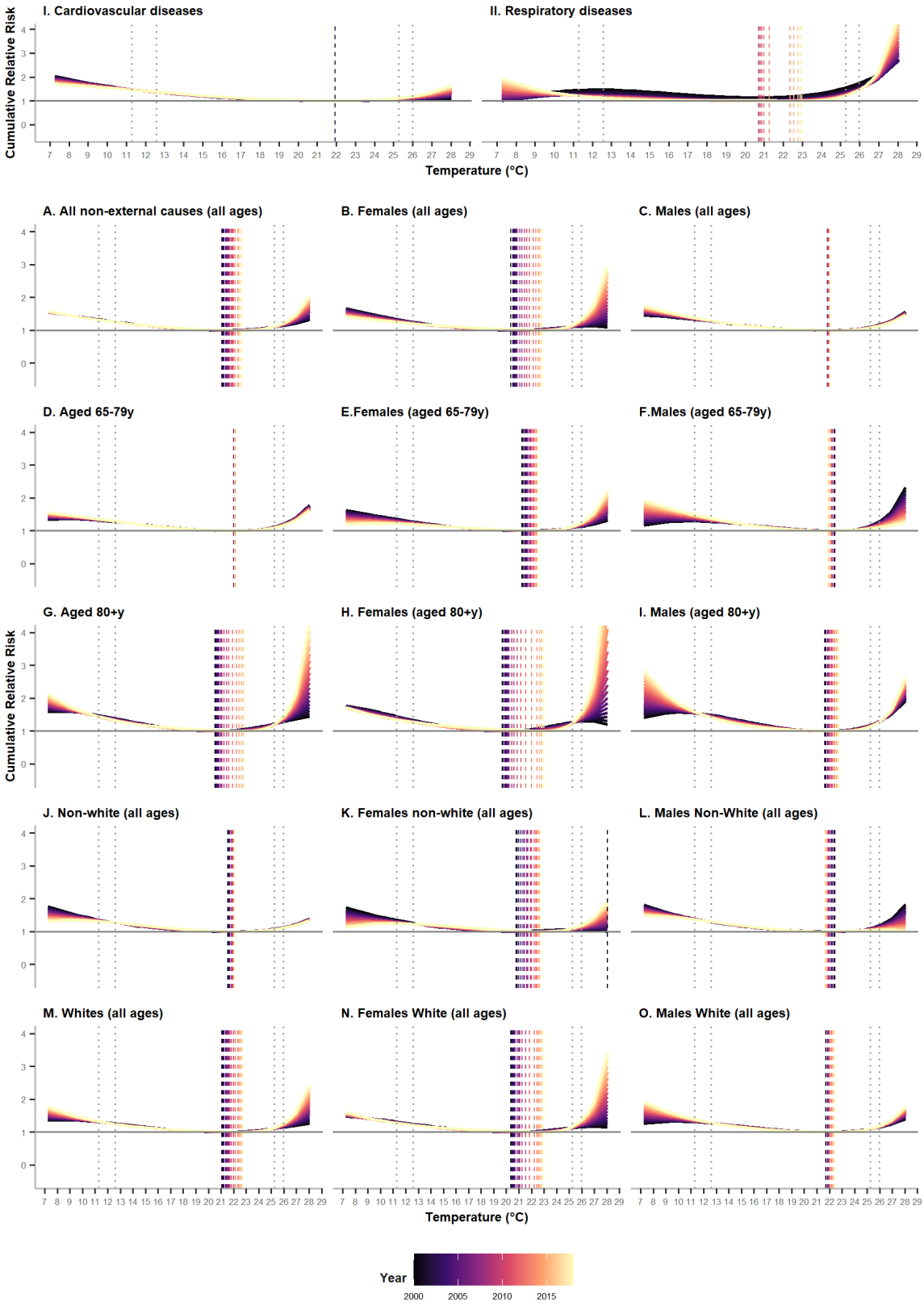


Figure 3.9: **Annual cRR by population group.** Annual cumulative relative risk (cRR, solid line) for the temperature-mortality association for (I) cardiovascular diseases (CVD) mortality and (II) respiratory diseases mortality, and for (A-S) all non-external causes of deaths by age, sex, ethnic groups, and combination of those. The annual MMT shown as a vertical dashed line. The 1^{st},10^{th}, $90^{th}$and 99^{th}percentiles of the temperature distribution are shown as dotted vertical lines.

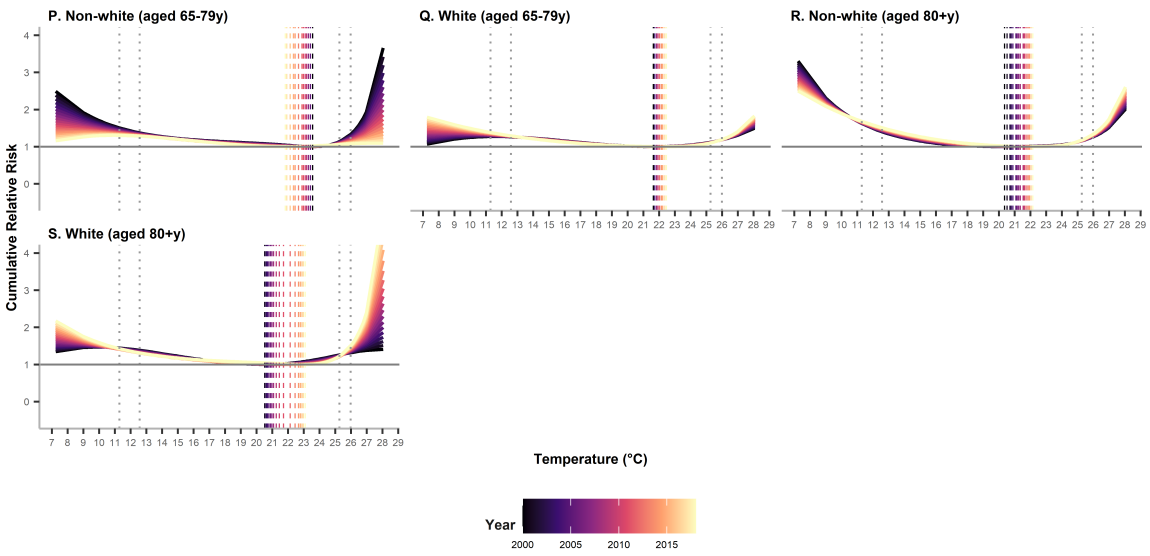


Figure 3.9: (cont.) Annual cumulative relative risk (cRR, solid line) for the temperature-mortality association for (I) cardiovascular diseases (CVD) mortality and (II) respiratory diseases mortality, and for (A-S) all non-external causes of deaths by age, sex, ethnic groups, and combination of those. The annual MMT shown as a vertical dashed line. The 1^{st},10^{th}, $90^{th}$and 99^{th}percentiles of the temperature distribution are shown as dotted vertical lines.

Table 3.8: cRR associated to extreme and moderate cold and heat by population group with 95%CI. Extreme/moderate heat and cold defined as the $99^{th}/90^{th}$and $1^{st}/10^{th}$ percentile of the annual temperature distribution estimated using annual group-specific MMT as reference. Last column, p-value of the Wald test for interaction (*, p-value <0.05).

			Extreme cold	Moderate cold	Moderate heat	Extreme heat	
Population group		Year	cRR(95% CI)	cRR(95% CI)	cRR(95% CI)	cRR(95% CI)	p-value
Overall		2000	1.33 (1.27;1.39)	1.14 (1.11;1.17)	1.04 (1.01;1.07)	1.14 (1.06;1.23)	0.00*
		2018	1.31 (1.25;1.38)	1.14 (1.11;1.18)	1.01 (1.00;1.03)	1.22 (1.16;1.28)	
Gender	Females	2000	1.34 (1.26;1.43)	1.12 (1.08;1.16)	1.06 (1.02;1.11)	1.12 (1.01;1.24)	0.01*
		2018	1.25 (1.16;1.34)	1.13 (1.08;1.18)	1.02 (1.00;1.03)	1.33 (1.24;1.42)	
	Males	2000	1.33 (1.25;1.40)	1.16 (1.12;1.20)	1.03 (1.00;1.06)	1.17 (1.06;1.29)	0.00*
		2018	1.38 (1.30;1.47)	1.16 (1.11;1.20)	1.02 (1.00;1.04)	1.13 (1.07;1.20)	
Age group	65-75 y	2000	1.32 (1.22;1.42)	1.17 (1.12;1.23)	1.03 (0.99;1.06)	1.21 (1.06;1.38)	0.51
		2018	1.35 (1.25;1.47)	1.17 (1.11;1.23)	1.02 (0.99;1.04)	1.16 (1.08;1.25)	
	≥80 y	2000	1.48 (1.36;1.61)	1.20 (1.14;1.27)	1.12 (1.05;1.18)	1.29 (1.12;1.49)	0.00*
		2018	1.45 (1.33;1.57)	1.18 (1.12;1.24)	1.02 (1.00;1.03)	1.46 (1.34;1.58)	
Ethnic group	White	2000	1.31 (1.24;1.37)	1.15 (1.11;1.19)	1.05 (1.01;1.08)	1.14 (1.05;1.24)	0.00*
		2018	1.33 (1.26;1.41)	1.13 (1.09;1.17)	1.01 (1.00;1.02)	1.27 (1.20;1.34)	
	Non-white	2000	1.36 (1.23;1.50)	1.14 (1.07;1.21)	1.02 (0.97;1.08)	1.13 (0.95;1.33)	0.04*
		2018	1.31 (1.19;1.44)	1.18 (1.11;1.25)	1.02 (0.99;1.05)	1.11 (1.02;1.21)	
Gender and Age group	Females 65-79y	2000	1.35 (1.22;1.51)	1.15 (1.07;1.23)	1.03 (0.97;1.10)	1.13 (0.94;1.35)	0.45
		2018	1.21 (1.07;1.36)	1.13 (1.05;1.21)	1.02 (0.99;1.05)	1.25 (1.12;1.39)	
	Females ≥80y	2000	1.45 (1.30;1.62)	1.15 (1.09;1.22)	1.18 (1.08;1.28)	1.30 (1.09;1.56)	0.02*
		2018	1.35 (1.21;1.50)	1.17 (1.09;1.25)	1.02 (1.01;1.04)	1.58 (1.42;1.75)	
	Males 65-79y	2000	1.29 (1.16;1.43)	1.20 (1.12;1.28)	1.02 (0.99;1.06)	1.28 (1.08;1.53)	0.50
		2018	1.49 (1.34;1.65)	1.20 (1.13;1.28)	1.02 (0.98;1.06)	1.10 (1.00;1.21)	
	Males ≥80y	2000	1.55 (1.36;1.76)	1.31 (1.21;1.43)	1.06 (0.98;1.13)	1.30 (1.03;1.63)	0.10
		2018	1.60 (1.41;1.81)	1.19 (1.1;1.29)	1.02 (0.99;1.04)	1.29 (1.14;1.46)	
Gender and Ethnic group	Females non-white	2000	1.35 (0.40;4.62)	1.13 (0.32;3.98)	1.05 (0.29;3.87)	1.07 (0.38;2.99)	0.47
		2018	1.25 (1.08;1.44)	1.18 (1.08;1.29)	1.01 (0.99;1.04)	1.20 (1.04;1.38)	
	Females white	2000	1.32 (1.22;1.41)	1.13 (1.08;1.18)	1.08 (1.02;1.13)	1.15 (1.02;1.30)	0.06
		2018	1.27 (1.18;1.38)	1.12 (1.07;1.18)	1.02 (1.00;1.03)	1.37 (1.27;1.48)	
	Males non-white	2000	1.39 (1.21;1.60)	1.16 (1.06;1.27)	1.02 (0.97;1.06)	1.19 (0.94;1.50)	0.18
		2018	1.37 (1.21;1.54)	1.18 (1.10;1.27)	1.02 (0.98;1.08)	1.05 (0.94;1.17)	
	Males white	2000	1.31 (1.22;1.39)	1.18 (1.13;1.23)	1.03 (0.99;1.06)	1.14 (1.02;1.28)	0.00**
		2018	1.39 (1.29;1.49)	1.14 (1.09;1.19)	1.01 (1.00;1.03)	1.18 (1.10;1.26)	
Ethnic and Age group	Non-white aged 65-79y	2000	1.51 (1.21;1.88)	1.22 (1.06;1.40)	1.00 (0.99;1.01)	1.33 (0.94;1.89)	0.75
		2018	1.30 (1.11;1.54)	1.20 (1.09;1.32)	1.02 (0.96;1.09)	1.08 (0.92;1.26)	
	Non-white aged ≥80y	2000	1.62 (1.26;2.09)	1.14 (0.98;1.33)	1.05 (0.87;1.26)	1.26 (0.82;1.94)	0.39
		2018	1.68 (1.37;2.07)	1.26 (1.11;1.43)	1.03 (0.96;1.10)	1.33 (1.08;1.65)	
	White aged 65-79y	2000	1.27 (1.17;1.38)	1.19 (1.13;1.26)	1.04 (0.99;1.08)	1.18 (1.03;1.36)	0.28
		2018	1.38 (1.26;1.52)	1.16 (1.09;1.22)	1.02 (0.99;1.04)	1.19 (1.09;1.30)	
	White aged ≥80y	2000	1.44 (1.32;1.58)	1.22 (1.15;1.29)	1.13 (1.06;1.21)	1.31 (1.12;1.54)	0.00*
		2018	1.43 (1.30;1.57)	1.17 (1.10;1.24)	1.02 (1.00;1.03)	1.50 (1.37;1.65)	

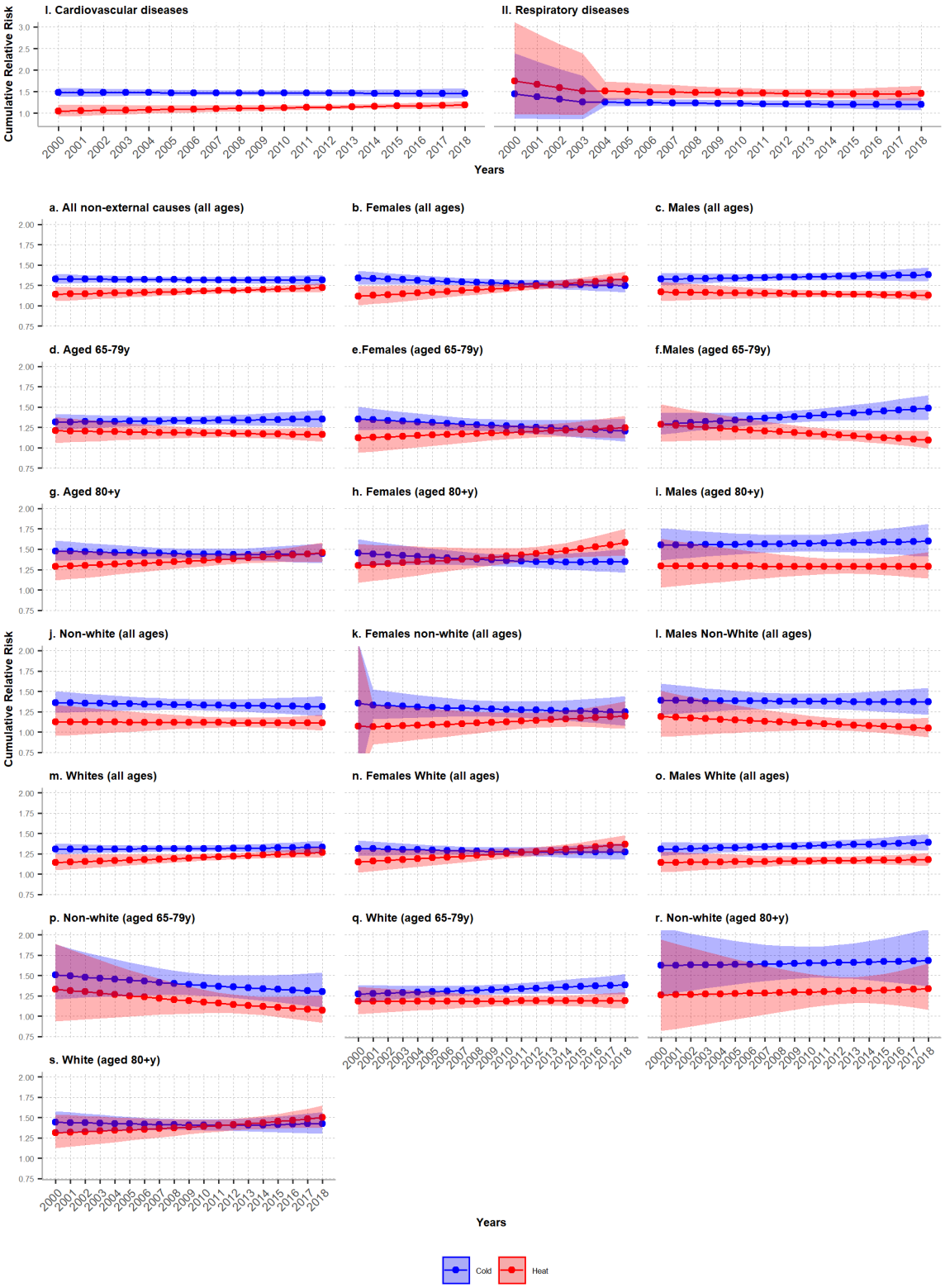


Figure 3.10: **Annual cumulative relative risk (cRR) by population group.** Annual cumulative relative risk (cRR) associated to extreme heat (red) and extreme cold (blue) for (I) cardiovascular diseases (CVD) mortality and (II) respiratory diseases mortality, and for (A-S) all non-external causes of deaths by age, sex, ethnic groups, and combination of those. The 95% CI are shown shaded in the same colour. Extreme heat and cold defined as the 99^{th} and 1^{st} percentile of the annual temperature distribution estimated using the annual specific MMT as the reference value, respectively.

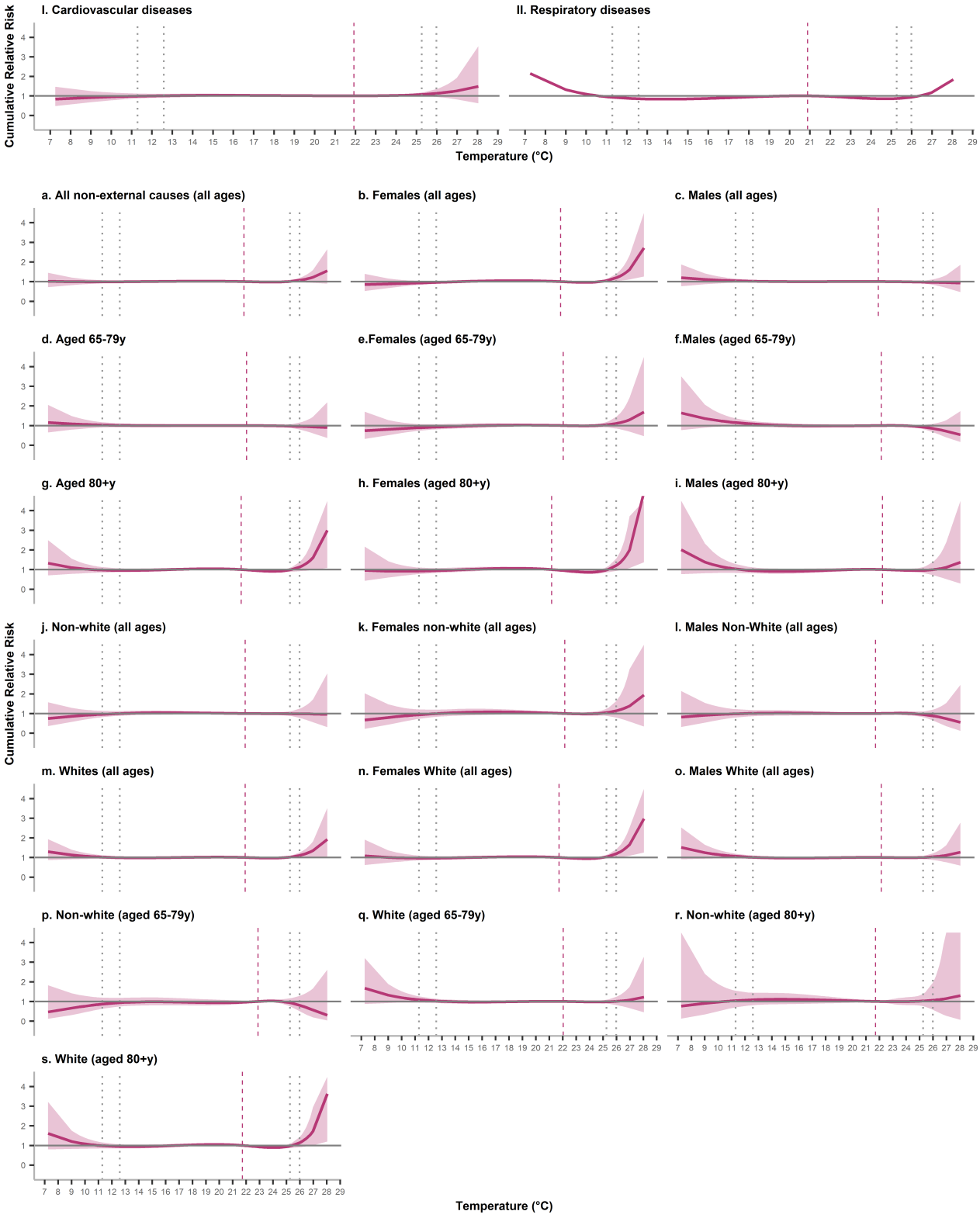


Figure 3.11: **Visual representation of the interaction terms by population group.** Cumulative relative risk (solid line) and 95% CI by cause of death, age, gender, and ethnic group. The MMT (dashed), and the 1^{st},10^{th}, 90^{th}, and 99^{th} percentiles of the temperature distribution (dotted) are shown as vertical lines. The curves were re-scaled to the MMT from the model without interaction (interpreted as the average across the whole study period)

3.3.2.3 Is climate driving the changes?

Figures 3.12, 3.13, and 3.14 provide a visual representation of the association of annual MMT, and extreme cold and heat with AMT, and annual extreme cold and heat, respectively. P-values of trend are provided in each plot together with the regression coefficient and Pearson correlation.

On the one hand, the MMT-AMT association was not statistically significant in any of the population group studied (Figure 3.12). These results seem to suggests that the MMT fluctuates over time independently of the annual mean temperature experienced, and thus, that there is no climate driven adaptation when it comes to MMT.

On the other hand, the expected association in the abscence of adaptation for cold would be negative (i.e., risk higher in low temperatures) and for heat to be positive (i.e., risk higher for high temperatures). In this case, if adaptation is present, I would expect to find a small to no association to extreme temperature, meaning that the risk is independent of the temperature experienced. A full reversion of the association would point to even stronger independence of adaptation from the experienced temperature and the presence of other external drivers with a bigger influence. Here, I found that the associations between cRR and temperature for extreme cold (Figure 3.13) and heat (Figure 3.14) were statistically significant in all cases, yet the coefficients remained rather low in most cases, suggesting a very weak association. I found non-white population aged 65-79 years old (-0.06), females aged 65-79 years old (-0.05) and females aged 80+ years old (-0.03) to have the strongest negative association with extreme cold (Figure 3.14). For heat, the groups with the strongest positive association with high temperatures were white females (0.06), females of all ages (0.06) and aged 80+ years old (0.05), population aged 80+years old (0.04), females non-white (0.04) and white population of all ages (0.03)(Figure 3.13). This suggest that there is still a strong correlation between the experienced temperature and the observed risk, suggesting a lack or poor adaptation. For the other groups, the association was weak, suggesting a slight decoupling between the extreme temperatures experienced suggesting some degree of adaptation.

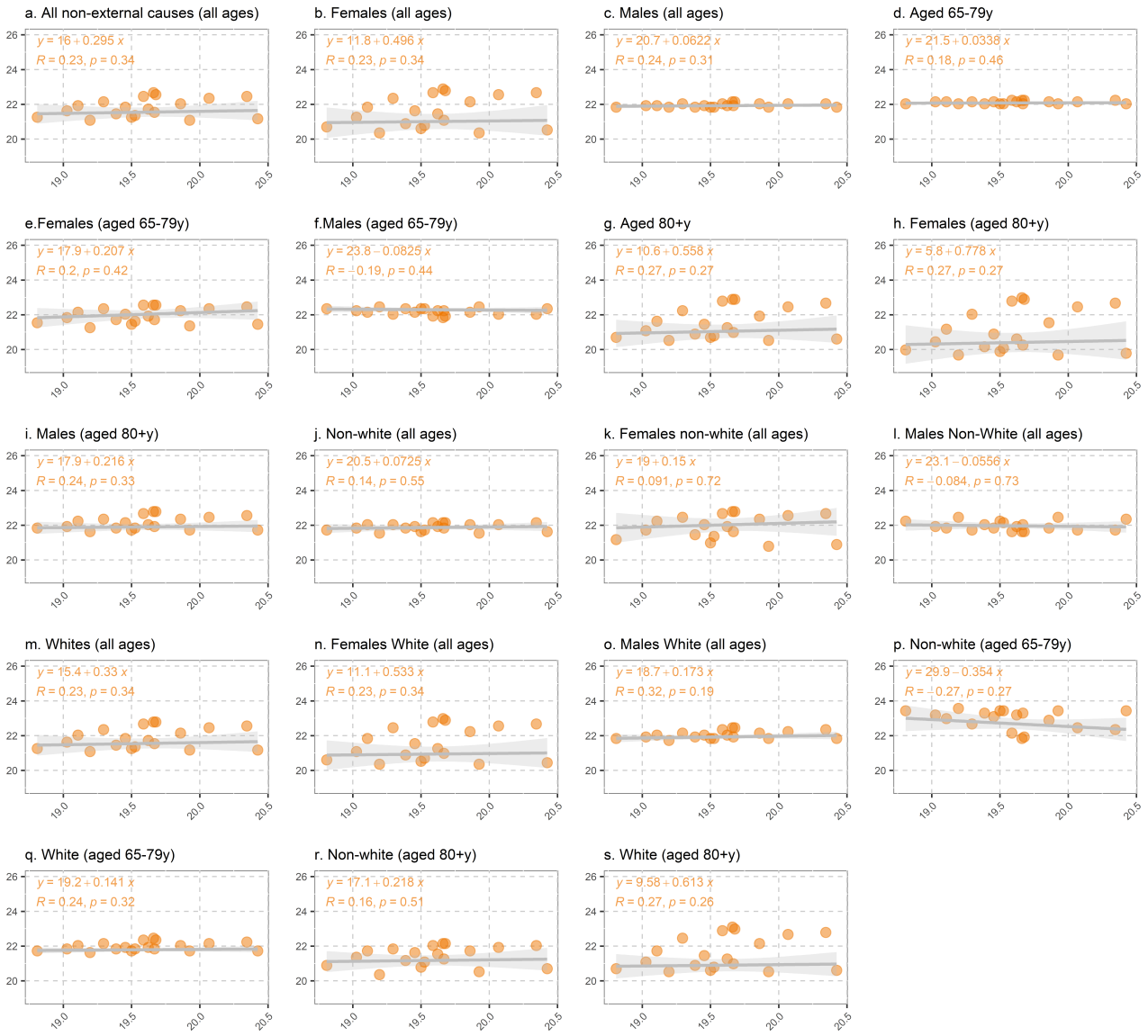


Figure 3.12: **Association between MMT and annual mean temperature.** On the y-axis, the annual MMT. On the x-axis, the annual mean temperature (AMT). Plots or each population group. Plots shown by population group. Regression line with 95% confidence intervals are shown as a solid grey line and a light grey shading; regression coefficients, Pearson correlation coefficient and p-value shown in orange.

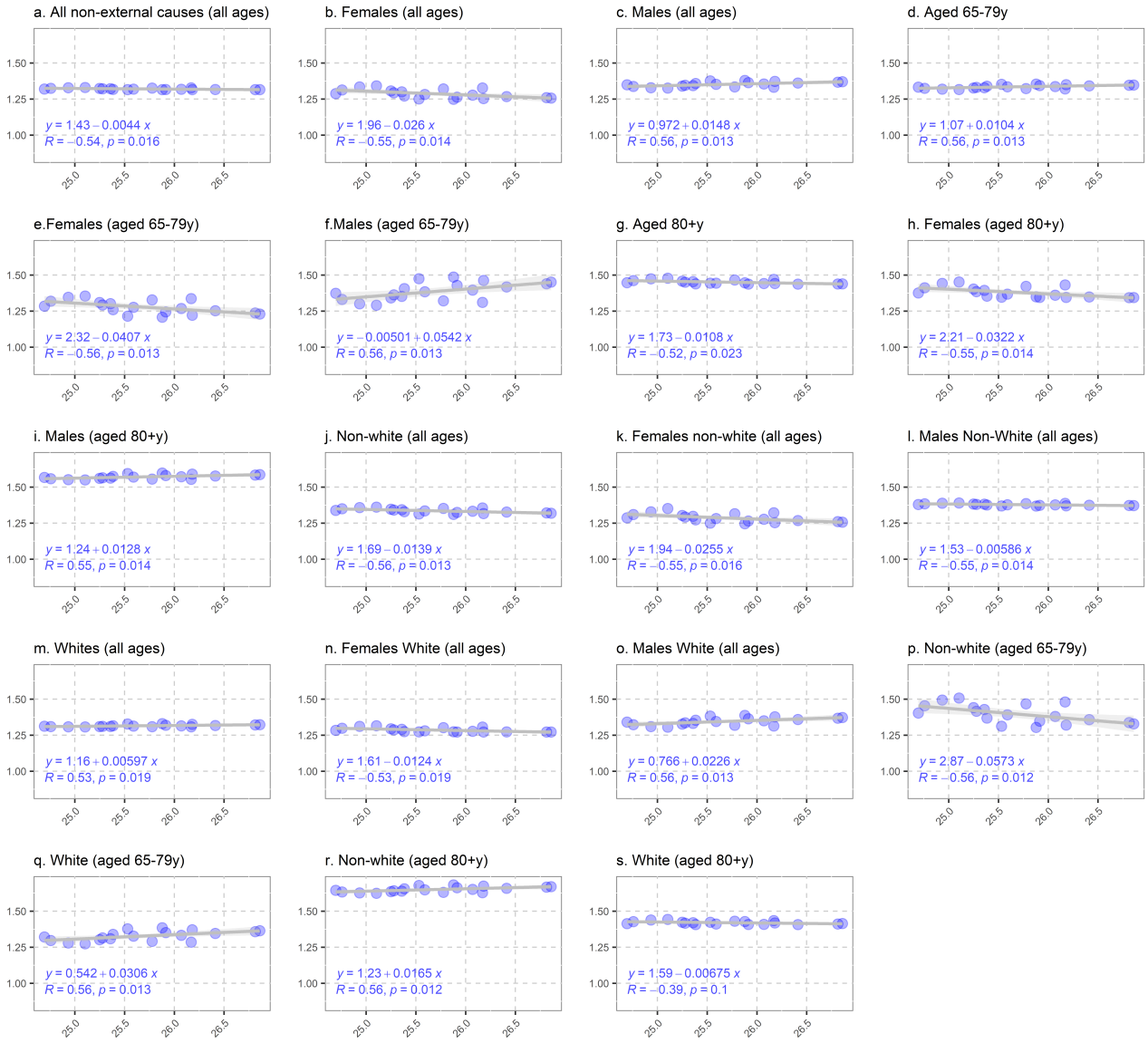


Figure 3.13: **Association between cRR for extreme cold[a] and annual extreme cold.** Y-axis, cRR associated to extreme cold defined as the cRR associated to the 1^{st} percentile of the annual temperature distribution. X-axis, the 1^{st} percentile of the annual temperature distribution. Plots shown by population group. Regression line and coefficients, Pearson correlation coefficient and p-value in blue. [a]Extreme cold: cRR at the 1^{st} percentile vs. MMT.

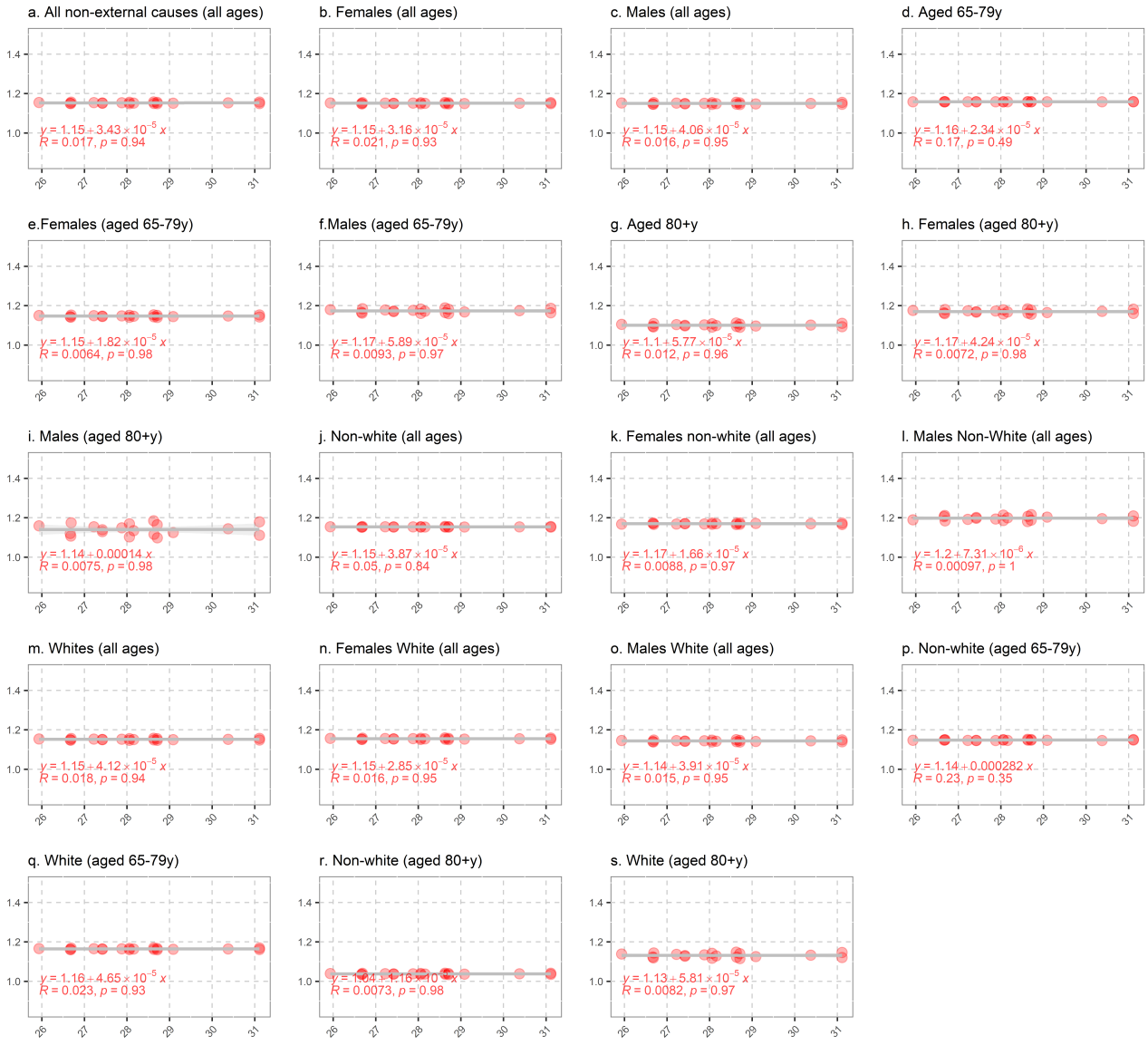


Figure 3.14: **Association between cRR for extreme heat[a] and annual extreme heat.** Y-axis, the cRR associated to extreme heat, defined as the cRR associated to the 99^{th} Percentile of the annual temperature distribution. X-axis, the 99^{th} percentile of annual temperature distribution. Plots shown by population group. Regression line and coefficients, Pearson correlation coefficients and p-value shown in red. [a]Extreme heat: cRR at the 99^{th} percentile vs. MMT.

3.3.3 Sensitivity analyses

3.3.3.1 Model parametrization

Table 3.4 shows the model parametrization used for each sensitivity analysis and the associated model performances, assessed through the qAIC score, for models with and without interaction term. The main model, which has the parameter configuration that resulted in the best performance ($qAIC_{noint}$ = 66501.7; $qAIC_{int}$ = 68362.2) is shaded. Appendix Figure 6.4 shows the MMT for each sensitivity analysis for all non-external causes of death. The selection of exposure and lag basis functions (Appendix Figure 6.4 A) and the location of knots (Appendix Figure 6.4 B) affected some of the MMTs, with some parametrizations yielding higher values and some others lower. In most cases, the estimates used in sensitivity analyses had larger uncertainty than the main model parametrization. They did not they change the direction of the temporal trend in any case. The number of df used for seasonality (Appendix Figure 6.4 D) only slightly affected the annual MMT, returning systematically lower values, and thus, not affecting the temporal trends. For the maximum lag parameter (Appendix Figure 6.4 C), there were substantial differences in the mean annual MMT; using shorter maximum lag periods (15 days) resulted in lower annual MMT yet larger variability, especially in later years; whereas, using larger maximum lag periods resulted in attenuated trends of MMT with large variability in the early years. In all tested scenarios, the trend continued to be positive over time. The 21 days maximum lag window was selected as it had the lowest uncertainty (i.e., smaller confidence intervals). The annual cRR estimates obtained from each sensitivity model, showed a similar behavior to those reported for the main model (Appendix Figure 6.6 and 6.5).

3.3.3.2 Relative humidity and air pollution

For RH, the best fitting parametrization was the one that used a 2-day moving average of the daily mean RH modelled as a natural cubic spline with 3 degrees of freedom. The use of other RH proxies had virtually no impact on the annual MMT estimates (Appendix Figure 6.4 E). The fluctuation in MMT estimates due to different model parametrizations was small, suggesting the model is robust. Similar behaviour observed for PM_{10}.

3.3.3.3 Gridded dataset

When comparing the ERA5-Land population-weighted daily mean temperature with the IAG-USP daily mean temperature, I found a high correlation (Pearson correlation coefficient: 0.97, p-value$<2.2e^{-16}$, Figure 3.15 A). The Root Mean Square Error (RMSE) was 0.89°C. There is also good alignment between the time series of both data sources (Figure 3.15C), indicating that the fluctuations over time are similar for both data sources. Nevertheless, when comparing the temperature distribution of both

temperature sources (Figure 3.15 B), we can see that the ERA5-Land estimates slightly underestimate extreme heat temperatures. Figure 3.16 and Table 3.9 compare the cumulative relative risk (cRR) curves and MMT from using IAG-USP and ERA5-Land with good alignment between both lines. The major difference lies in the tail for high temperatures.

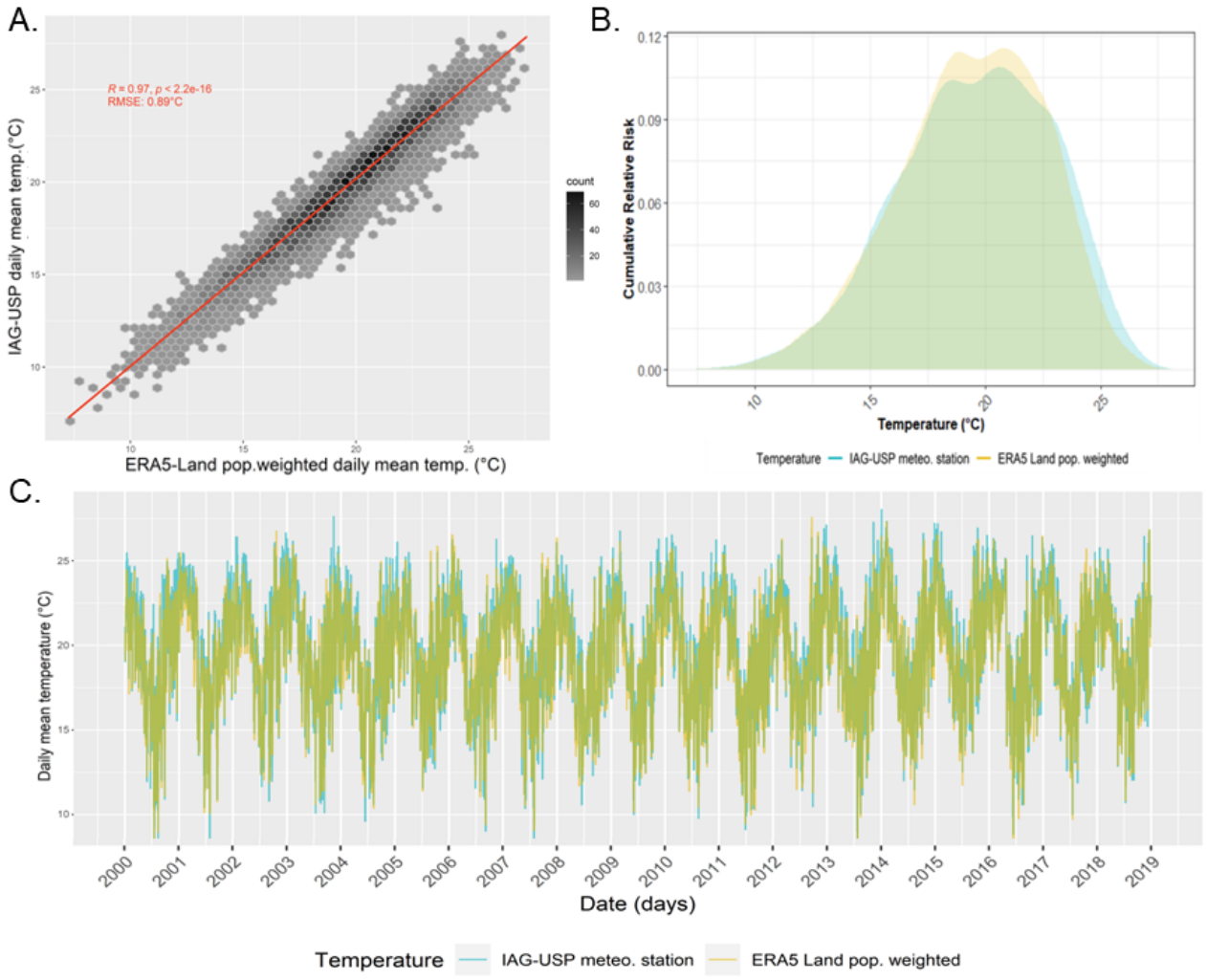


Figure 3.15: **ERA5-Land and IAG-USP temperature comparisons.** (A) Scatter plot between ERA5-Land population-weighted (x-axis) and IAG-USP meteorological station (y-axis) daily mean temperature. The regression line, root mean square error (RMSE), and Pearson correlation coefficient and p-value in red. (B) Histogram and (C) time series plots of daily mean temperature from the IAG-USP meteorological station (green) and population-weighted ERA5-Land estimates (yellow).

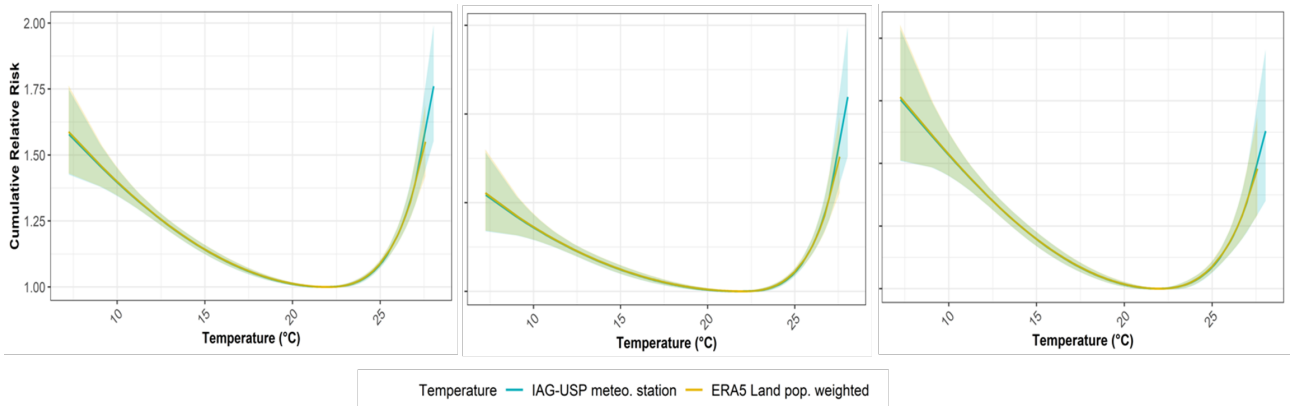


Figure 3.16: **ERA5-Land and IAG-USP temperature-mortality association.** From left to right, overall cumulative exposure-response associations for all non-external causes of death in the state of São Paulo among all population, females and males, respectively. The IAG-USP meteorological station observations in blue and the population-weighted ERA5 Land in yellow, with 95% confidence intervals (95%CI) shaded.

Table 3.9: **ERA5-Land and IAG-USP temperature-mortality association: cRR and MMT.** Cumulative relative risks (cRRs) at the 1^{st} and the 99^{th} percentile, and minimum mortality temperature (MMT) for all non-external deaths overall, and among females and males, using IAG-USP and ERA5-Land temperature data. The 95% confidence intervals (95% CI) are also provided.

		MMT(95%CI)	cRR cold (95%CI)[a]	cRR heat (95%CI)[b]
Overall	IAG-USP	21.83 (21.45;22.46)	1.46 (1.38;1.54)	1.19 (1.16;1.23)
	ERA5-Land	21.99 (21.45;22.37)	1.45 (1.38;1.53)	1.14 (1.12;1.16)
Females	IAG-USP	21.83 (21.08; 22.56)	1.42 (1.31;1.54)	1.25 (1.21;1.30)
	ERA5-Land	21.99 (21.19;22.50)	1.42 (1.32;1.53)	1.18 (1.15;1.21)
Males	IAG-USP	21.93 (21.36;22.56)	1.49 (1.39;1.60)	1.15 (1.11;1.19)
	ERA5-Land	21.99 (21.38;22.59)	1.48 (1.39;1.59)	1.11 (1.08;1.14)

[a] cRR cold: $1^{s}t$percentile vs. MMT; [b] cRR heat: $99^{t}h$percentile vs. MMT

3.4 Discussion

This study investigated trends in the temperature-mortality association by population groups in São Paulo, Brazil, between 2000 and 2018. Below, I describe in detail each the main findings in the context of existing literature (Section 3.4.2). I also discuss the role that the temperature data source and the confounders may have played in the observed findings in this specific context (Section 3.4.3). Finally, I provide an overview of the public health implications and strengths and limitations of the study.

3.4.1 Temperature-mortality association in São Paulo

Overall, when analyzing all non-external causes of death across all ages during the entire study period, a steep rise in the heat-mortality association was observed over a short temperature range deviating from the MMT ($temp_{\text{MMT-ExtremeHeat}}=$), indicating high susceptibility to temperatures above the MMT. For cold, the rate of increase in risk per unit of was much lower. However, due to the left-skewed

temperature distribution, the cut-off value for mild and extreme cold was farther from the MM (temp $_{\text{MMT-ExtremeCold}}$=), resulting in risk levels comparable or higher than those for heat. These differences likely stem from differences in the level of acclimatization, adaptation responses, and susceptibility factors to heat and cold. Similar shapes were described by Gasparrini et al. 2015 [3]. In their multi-city study including 384 locations worldwide and over 74 million deaths between 1985-2012, researchers found substantial regional variations in the temperature-mortality association. Interestingly, tropical and subtropical locations, including São Paulo, exhibited a steeper cold risk function over a narrower temperature range than other regions, though still less steep than the heat-related risk function. This suggest a worse adaptation to cold compared to temperate regions.

In addition to the shape of the association, it is important to consider the underlying temperature distribution and the cut-off points used to define the risks. In estimating the cold-related relative risk (cRR), it is essential to identify reference points for defining mild or extreme temperatures. A common method involves utilizing the overall temperature distribution for the specific study area and period. Here, I used the 10^{th} and 1^{st}, and the 90^{th} and 99^{th} percentiles of the overall temperature distribution to define mild and extreme cold and warm temperatures, respectively. In São Paulo, the temperature distribution exhibited a leftward skew (negative skewness), indicating a prolonged tail for cold temperatures. Consequently, the cut-off points for estimating cRR in cold conditions displayed greater variability (SD=, range=) than those for heat (SD=, range=), featuring values that deviated significantly from the MMT (temp$_{\text{MMT-maxTemp}}$= ; temp $_{\text{MMT-minTemp}}$=). As a result, the risk estimates for cold are likely to capture more extreme situations with a larger effect magnitude but also entail greater variability and uncertainty.

3.4.2 Trends in temperature-mortality association: adaptation

Three main findings arose from this research. Firstly, that the temperature-mortality relationship is a dynamic process that varies over time. When looking at the overall population, I observed a 1.7°C increase in the MMT for all non-external causes of death, which is in line with findings from previous studies in France[174, 233], Spain[234] , Sweden [235], Austria [101], and the Netherlands[94]. Nonetheless, this is in contrast with the findings from Vicedo-Cabrera et al.,'s study[103], where the MMT reduced by -0.4°C from 1997 to 2011. Differences in the local context, the period under study and modelling choices may explain the conflicting results.

Overall, I also observed a slight increase in the heat-related mortality risk despite the annual increase of MMT, whilst cRR for cold-related mortality remained almost unchanged. This contrasts the most of the evidence suggesting a decrease in population's susceptibility to heat across a number of different settings [41]. In Vicedo-Cabrera et al.,'s study, which included 18 Brazilian cities[103] amongst which, São Paulo, the authors reported a decrease in the cRR for heat from 1.095 (95%CI:1.038;1.155)

to 1.064 (95%CI: 1.015;1.117), the cRR of cold decreased from 1.26 (95%CI: 1.196;1.347) to 1.222 (95%CI: 1.151;1.298), between 1997 and 2011. However, the findings were not statistically significant (Wald test: 0.471). When comparing to other countries, my results for heat contrasts with those reported in Spain [84] or in multi-city studies [40, 95] with some exceptions, such as the increase in heat-related mortality risk reported for Australia in Gasparrini et al., 2015[40]. Findings for cold are limited and contradicting [30, 40, 84, 95, 103].

Secondly, the trends in MMT differed by population group with a general of increase of ¿1°C for most of the groups, except for males aged 65-79 years old, non-white males, and won-white population aged 65-79 years old, whom showed a reduction in the MMT. Similarly, the trends in the cRR for heat and cold varied substantially across groups. In general, cRR for heat increased among females whilst it decreased among males; the opposite occurred for cold-related cRR. Among the eldest group, the cRR for heat increased regardless of gender and ethnicity; whilst the cold-related cRR remained more stable. Individuals with a non-white ethnic background showed more stable estimates than the white-ethnic background group, yet the associated uncertainty was higher (due to smaller numbers). Differences in vulnerability to temperature effects by gender, sex and ethnic groups are well-documented [13] and can be seen reflected in the risk estimates for the overall period herein reported. However, less studies have addressed temporal trends by vulnerable groups [41], as I did here. Achebak et al., 2018 [84] explored trends in the risk of CVD, respiratory diseases and combined mortality by sex group and found that despite there were slight difference in the magnitude of the adaptation, all groups experienced a decrease in cRR for extreme heat, defined by the authors as the 99^{th} percentile of the overall temperature distribution. Overall differences in the temperature-mortality risk by sex, age and ethnic groups, are often associated to behavioural, technological, occupational or socioeconomic differences (see Chapter 2 Section 2.4). These factors are likely to vary over time affecting differently each group and so, influencing their adaptive capacity. Such factors may include: improvements in the life expectancy at birth (70.1 years in 2000; 75.7 years in 2018) [23]; increase in health care coverage (from 7.8% to 58.5% between 2000 and 2016 [236]), adoption of healthier lifestyles, e.g., decreased smoking prevalence (from 35.3% in 1989 to 11.3% prevalence among adults $\geq$20 years old in 2017 [237]), among others. Socioeconomic status is also a known effect modifier of the temperature-mortality association [179, 215]. In 2014, Brazil experience an economic recession that caused a doubling in the unemployment rates in São Paulo which most likely modified the exposure pathways to temperature [238]. The analyses by educational level aimed at exploring the impact of job market and socioeconomic status, but the poor quality of the data hindered its use. Finally, I did not observe a significant change in the AMT nor in extreme temperatures. Nonetheless, I cannot rule out the role of climate-driven physiological and behavioural adaptation, e.g., uptake of air conditioning [85, 239] or raised awareness of weather-related health impacts.

Finally, the changes observed seemed to be only partially coupled to changes in temperature indicators, suggesting the presence of non-climatic adaptation drivers. This is in line with findings reported in Vicedo-Cabrera et al.,'s 2017 study[103]. Adaptation involves the capability to change one's risk profile by adjusting either exposure or vulnerability. This adaptive capacity is frequently interconnected with physiological, behavioural, infrastructural, and technological factors, leading to variations across different times and locations. Epidemiological studies often do not distinguish between types of adaptation and instead, treat adaptation as a residual explanation for observed changes in mortality effects over time, as done in my study. The observed effect is a combination of different mechanisms operating at distinct timescales and thus, disentangling the mechanisms behind the observed trends is nearly impossible. The decoupling of MMT and risk of death for cold observed in my study, seems to suggest the presence of factors other than temperature changes. The heterogeneous trends found across population groups further support the presence of non-climatic drivers of change as different groups are likely to adopt, have access to and use different adaptive behaviours, technology and infrastructure, which may explain the heterogeneity in the observed cRR and MMT trends. Of all adaptation measures, air conditioning and heat warning systems have been studied the most [240]. Interestingly, Viggotti et al. 2004 [241] investigated whether birthplace had an impact on the mortality–temperature relationships, using data from Milan. Namely, the authors found that mortality risks differ by birthplace, regardless of the place of residence, and that heat tolerance in adult life could be modulated by outdoor temperature experienced early in life. Despite birthplace information was available in this study, the quality of the data was not deemed enough to run such an analysis, yet it deserves further exploration.

3.4.3 Temperature and confounder data

From a physiological point the role of humidity seems to be obvious, however, in epidemiological studies, such association is less apparent. For example, in a multi-city study by Armstrong et al., 2019 [52] assessing 445 cities from 24 countries, the authors found that the effect of RH was almost negligible most settings. São Paulo is characterized by a really humid climate with little variability, as supported by my data (mean: 80%, SD: 8.5%), thus I did not expect RH to have a major role in this setting. When added to the model, it only improved slightly the performance and its coefficient was small and not significant, further supporting this hypothesis.

Some studies [36, 37] have suggested the use of re-analyses data, such as the ERA5-Land dataset, instead of a reference meteorological station. When compared to each other, ERA5-Land data and the IAG-USP meteorological data showed good alignment, which is not surprising as ground station measurements are an important component of the reanalysis datasets. My findings are in line with those from Misty et al., 2022 [36] who found good agreement between ERA5 daily temperature esti-

mates and weather stations using data from 612 cities across the world. It is also in line with evidence from Brazil using ERA5-Land data by Araújo et al., 2022[242]. In this paper, the authors report a mean of R^2 and RMSE across all stations of 0.92 and 0.63°C, respectively, indicating good quality of fit and agreement between ERA5-Land reanalysis and most automatic weather stations. Nevertheless, the ERA5-Land estimates seem to slightly underestimate the heat extremes. This could be a result of the larger spatial coverage of the data, which may dilute the extreme temperatures as it includes more peri-urban areas with high prevalence of greenspace. On the other hand, previous evidence shows that ERA5 (sister product of ERA5-Land) systematically underestimates the increasing frequencies of heat extremes [21]. This is likely to be a result of the fact that the input data used as ground truth only includes meteorological background stations and excludes urban and industrial meteorological stations. This could explain why, despite the correlation being high for most of the time series (Figure 3.15; Pearson correlation coefficient: 0.97), ERA5-Land has fewer extreme temperatures recorded for heat. Mistry et al., 2022 [36] estimated the correlation and goodness-of-fit of with city reference stations across 612 cities globally, and compared the temperature-related risk estimates derived from using each temperature source. They found a good alignment between the temperature estimates and similar non-optimal temperature-related risk estimates. However, sub-tropical and tropical cities showed a poorer correlation (São Paulo: 0.95; RMSE: 1.45), mainly due to underestimation of heat. This is in line with my findings and an important limitation of ERA5-Land products for my study, as I are particularly interested in temperature extremes.

The impact of temperature data source in temperature-mortality associations has been studied previously, and good agreement has been reported in most cases [36, 37]. Here, I compared the temperature-mortality association obtained using IAG-USP meteorological station data to that obtained using ERA5-Land population-weighted daily temperature. Despite the risk estimates and MMT compared overall, using IAG-USP showed a better performance for extreme heat, suggesting that they are better fit for this study. This is in line with the underestimation of heat reported above for ERA5-Land.

In short, reanalysis data stands as a solid alternative to monitoring stations when these are not available and for city-wide studies. Its global and continuous spatial and temporal coverage at a relatively high spatial resolution makes it an attractive dataset for epidemiological studies. However, they have been shown to suffer from mean bias, often underestimating heat extremes [21], as I also showed in my study. This makes its performance in urban areas, where phenomena like the urban heat island are in place, much poorer. Moreover, their coarse resolution, yet adequate to study large areas, it is not capable of capturing the micro-climates that exist in cites [15]. This calls for novel approaches capable of predicting temperature at a high spatial resolution without compromising the temporal resolution. This will be addressed in Chapter 4 of this thesis.

3.4.4 Public health implications

Overall, my findings suggest the need for local, up-to-date and population group-specific evidence to inform public health measures. One such measure is HHWS. These are short-term responsive measures aimed at minimizing the health impacts of extreme temperature activated upon reaching temperatures of risk for health. According to the WHO guidelines[32], the gold standard to define the temperature thresholds of activation is to consider setting-specific associated health outcome. Currently, São Paulo has a cold spell warning system for the homeless population only. It uses pre-defined temperature threshold based on daily minimum temperature (moderate risk: 13 °C; high risk: 10 °C)[243]. Based on my finding, at a daily mean temperature of 10°C, the general population experienced a 40% increase in mortality. These estimates are not specific for homeless population nor do they explore minimum temperatures and so, are likely underestimating the impact of cold extreme events. Even so, they suggest that São Paulo could substantially benefit from the development and implementation of a local HHWS which considers population groups, and temporal variations in the risk.

3.4.5 Strengths and novelty

My study has several strengths worth discussing. Firstly, it presents a robust and comprehensive investigation into the impacts of temperature on mortality over time focusing on a wide variety of population groups, aside from the often studied gender and sex categories. This approach highlights the significance of addressing vulnerability disparities and tailoring adaptation strategies to suit diverse needs. The robustness of the findings is strengthened by the large and diverse dataset, ensuring reliable representations of multiple population segments. Secondly, the use of a flexible modelling framework, the . This class of models surpass conventional linear models in capturing complex non-linear and delayed dependencies between temperature and mortality, while also allowing these to change over time (by year in my case). By accounting for the intricate temporal variations, this methodology enhances the accuracy and granularity of the analysis, providing valuable insights into the temporal trends of the mortality risk associated to temperature. Finally, my study takes a comprehensive approach to assess adaptation strategies. By employing multiple indicators, my research offers a holistic overview of temperature trends. This multifaceted evaluation enables policymakers and stakeholders to make informed decisions based on a more nuanced understanding of adaptation trends and their impacts on mortality outcomes.

3.4.6 Limitations

Despite the rigorously and comprehensiveness of the study, it is essential to acknowledge certain limitations that affect my study.

Firstly, the analyses were conducted using data from only one meteorological station which may have caused exposure misclassification due to spatial heterogeneity in the distribution of temperature. It is important to emphasize that this investigation was part of an initial exploratory phase to gain insights into temporal trends before committing to a more resource-intensive spatial analysis. The convenience and accessibility of data from a single meteorological station played a pivotal role. Moreover, it aligns with a common approach in epidemiological studies that prioritize in situ measurements as the gold standard. In situ measurements from meteorological stations provide real measurements of the temperature experienced by individuals on the ground with a high temporal resolution (minutes, hours, or days). This is why, despite their spatial representation and coverage is poor in the best cases, meteorological stations are still often regarded as the gold-standard in epidemiological studies and the most common proxy data used for population exposure to outdoor temperature [40]. Also, the selected study design places the focus on the temporal component over spatial granularity, given the primary objective of discerning temporal trends in temperature patterns across the entire state. Thus, any spatial granularity obtained from more complex temperature assessments would have been diluted in the analyses stage. Despite the acknowledged limitations, a comparison with a more spatially granular dataset, the ERA5-Land daily mean temperature dataset, was conducted to address potential bias concerns. The results indicated slight underestimation of extreme heat by ERA5-Land, with minimal differences consistent with previous studies. In short, this analysis constituted a logical first step toward a more comprehensive evaluation, informing subsequent decisions in the PhD. , including the development of a temperature model for high spatiotemporal resolution. Finally, a comparison across different exposure proxies, as done here, is an advisable approach to help in the interpretation of the results and understand the possibilities of bias.

Secondly, the current study utilized 19 years of data, which despite being consistent with prevailing research practices, may be insufficient to capture certain adaptive patterns. Adaptation is influenced by physiological, behavioral, infrastructural, and technological factors, all with different time frames of implementation, for example, infrastructure changes may take years or decades, while behavioral shifts may be visible in shorter time frames. Thus, there is no universal consensus on a minimum time period to assess adaptation, yet longer periods are favoured. Martinez-Solanas et al. 2019 [244], Heutel et al. 2021 [245], and Folkerts et al. 2020 [94], employed a similar time range with 20 to 25 years of data. Longer study periods are less common due to lack of data or poor data quality for earlier periods. Some examples of studies utilizing data ranging from 25 to 35 years were Gasparrini et al. 2016 [95] and Follos 2020 [234], while Ordanovich et al. 2023 [246] and Acheback et al. 2019 [89] exceeded 35 years.

Another significant limitation of this approach is the lack of information on specific adaptation mechanisms, hindering a comprehensive understanding of community adaptation dynamics. Of

all adaptation measures, air conditioning [240] and heat warning systems [244] have been studied the most studied. Data availability is often the main constrain in assessing adaptative mechanisms. his information gap potentially restricts the identification of specific factors influencing adaptation outcomes which could be pivotal for informing policymakers and stakeholders and directing their strategies. In light of this limitation, alternative study designs and approaches should be explored. For example, looking at birth place in an attempt to disentangle physiological adaptation to societal adaptation, as done by Vigotti et al. 2006 [241] or Evangelopoulos D et al. 2011 [247]. Despite birthplace information was available in this study, the quality of the data was not deemed enough to run such an analysis, yet it deserves further exploration.

Thirdly, there may be residual confounding in my modelling framework, mainly due to lack of data or the use of proxies which do not fully capture the confounder. My model relied solely on PM_{10} data for evaluating the impact of air pollution on mortality, as access and/or the quality of other air pollutants was insufficient. The unaddressed effect of changes in other air pollutants may contribute to confounding the results, necessitating future studies to consider a broader range of pollutants to gain a more comprehensive understanding of their influence on mortality risks. Furthermore, air pollution has been suggested in several occasions to have synergistic effects with temperature [248] or as a mediator [249]. Future studies should investigate this phenomenon in detail for São Paulo. Finally, there may be other confounders at play such as influenza epidemics, sunlight exposure, and atmospheric pressure, which were not accounted for in the study due to a lack of data.

Finally, a potential limitation arises from utilizing data from a single meteorological station, potentially obscuring spatial patterns in the distribution of mortality risk. The exclusion of spatial variations might smooth the risk estimates obtained, highlighting the need for accurate and highly resolved temperature datasets, which I address in the next Chapter.

3.5 Summary

In this chapter, I have provided evidence of disparities in the adaptation capacity across population groups and temperature ranges (cold *vs.* heat) in São Paulo. I did not find a clear association between the indicators of adaptation and different temperature indicators over the study period. I argue that non-climatic factors with different penetration rates in the population are likely to have influenced such patterns. This study represents a significant step towards understanding climate change adaptation in mortality risk across diverse population groups. By incorporating advanced modeling techniques and employing a multiple indicators to explore changes (i.e., MMT and cRR for extremes), my research offers valuable insights for policymakers and stakeholders. While acknowledging certain limitations, my findings support the need for regular monitoring of the temperature-mortality association by

population groups to suitably inform public health interventions.

Chapter 4

Statistical modelling of daily mean temperature at high spatial resolution

4.1 Overview and objectives

More than half of the world's population lives in cities, a proportion expected to increase to 68.4% by 2050 [22]. Urbanization is considered one of the most dramatic and influential forms of land conversion of our times. In most cases, it involves the modification of previously natural or agricultural land into urban uses built on impervious surface, often characterized by areas with high building and population density. Although urbanization can take various shapes and forms (e.g., different building density, city layout or growth rate), its impact on the local environmental and climatic characteristics is undeniable [250–252]. For example, urbanisation affects the local surface energy balance, which alters the local surface and air temperature, creating micro-climates, of particular interest in this context. One of such phenomena, and probably a signature mark of the impacts of urbanization on climate, is the UHI effect by which cities systematically experience higher surface and air temperature than the surrounding rural areas [71].

Most studies investigating the temperature-mortality association use temperature recordings from one meteorological station or the arithmetic average across all available stations [3, 112]. This approach assumes temperature is distributed homogeneously across the study area. From studies on urban heat island, remote sensing data and knowledge from atmospheric and physical science, it is clear that this assumption does not hold [164]. Overlooking the spatial component of temperature exposure may bias epidemiological findings [253]. According to a recent systematic review [254], there is solid evidence suggesting that city micro-climates are strong effect modifiers of the temperature-mortality association. This is particularly of interest when informing public health interventions or tools such as HHWS and HHAP. In this context, spatially and temporally resolved temperature data are indispensable for providing reliable and accurate evidence of the impacts of temperature on population's health. In

practice, access to this data is challenging or not available, in particular for LMIC [20].

With over 11.7 million inhabitants over an area of 1,521.1km^2, the municipality of São Paulo is the largest urban agglomeration and most populous urban area in Brazil[22]. It has a varied urban fabric with a large spatial heterogeneity in the building type, building density and layout, distribution of green and blue spaces and population density, just to name but a few aspects. This heterogeneous urban built most certainly leads to spatial variability in the temperature distribution and so, potential exposure misclassification. This is already seen in studies assessing the spatial and temporal distribution of the UHI across the city, which has been changing over time as the city evolved [24]. Highly resolved temperature models can help to identify the number of people exposed, whether exposure is inequitably distributed within a city, and be used to assess the health impacts in more detail. The above are all key to inform urban planners and local governments, policymakers and stakeholders in developing area and group-targeted interventions to reduce the exposure and minimize the negative health outcomes of non-optimal temperatures.

In this chapter, I estimated daily mean temperature over the municipality of São Paulo at 500m^2, between 2015 and 2019. Of the methods discussed in Chapter 2, I selected a RF algorithm. Below, I first describe the modelling framework of choice and justify its selection. To gauge its effectiveness, I conducted a comparative analysis with a conventional algorithm, namely the Land Use Regression (LUR) or multi-linear regression model. Finally, I discuss the results in the context of the literature and present some potential future lines of research. The work presented in this chapter benefited from the knowledge and skills learned during my research visit at the European Space Agency (ESA) ϕ-lab in Frascati, Italy under the supervision of Dr. Rochelle Schneider. For a detailed description of the objectives and outputs of this research visit, refer to Appendix Section 6.7.3.

4.2 Data and Methods

4.2.1 Rationale for the Random Forest

The overarching objective of this modelling exercise was to generate a temperature dataset at high spatio-temporal resolution suitable to be used for epidemiological studies. The municipality of São Paulo is a really dense and heterogeneous urban area and so, in order to capture the spatial variability of temperature, I set the target spatial resolution to be 500m^2. To model the delayed and non-linear effects of temperature, I set the temporal resolution to daily.

With that in mind, of the approaches discussed in Chapter 2 Section 2.3.2.4, I discarded numerical climate models and their downscaled derivatives for their high computation expense and complexity. Regarding the reanalyses data, they were deemed to be of an insufficient spatial resolution to explore intra-city patterns, with ERA5-Land having the finest spatial resolution at 9km^2. Thus, the only

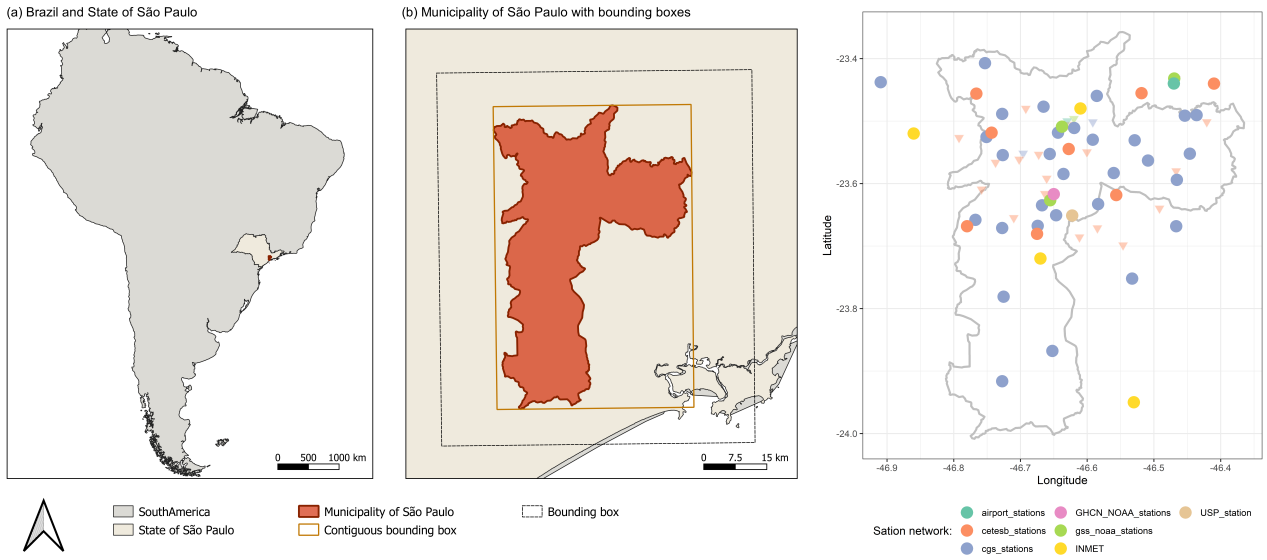
available alternative was to build and train a model using ground observed temperature data and covariates that could capture the spatial and temporal distribution of daily mean temperature. Fortunately, the municipality of São Paulo has a dense network of meteorological stations providing good quality daily temperature recordings. In view of this, regression and machine learning approaches were considered. Based on the strengths and limitations of the different available methods, which have been described in Chapter 2 Section 2.3.2.4, I decided on the RF algorithm as my modelling framework. Compared with other interpolation methods, such as LUR or STRK, the RF approach has several advantages. Firstly, RF is non-parametric, which means it makes no a priori assumptions about the relationship between the input features and the response. Thus, RF can capture complex associations between the input features and the dependent variable, including non-linear associations and interactions. Secondly, it can handle large number of input features. Compared to other machine learning algorithms, RFs are less sensitive to over-fitting and the presence of outliers, and provide useful measures of variable importance that allow the interpretation of the fitted model, such as the *feature importance ranking* which ranks features based on their weight in the predicting process. Naturally, RFs are not exempt of limitations. An important drawback of RF, is that it does not directly model the spatio-temporal autocorrelation, instead it relies on the input features to capture most of the variability, resulting in a negligible residual autocorrelation. This issue has been shown to be minor when using a large number of high quality and spatio-temporal resolved input features, as it is the case in this study with 43 input features at high spatial and temporal resolution and of good quality. Additionally, predictions from RFs are bound to the range of the observed data used during training, whereas spatial regression methods can extrapolate outside this range. This can be a strength when modelling responses that are physically constrained to a certain range. However, this is not the case for temperature. In order to investigate the extent of this issue, I compared the outputs of the RF to a traditional spatial regression models, the LUR. I also explored ways to address this challenge in Chapter 6 Section 6.5.1.4.

4.2.2 Study area

The study area for this modelling exercise is a rectangular bounding box (lon_{min}: -46.9559; lat_{min}: -24.0854; lon_{max}: -46.2226; lat_{max}: -23.2839, coordinate system: WGS84) around the municipality of São Paulo, Brazil. I extended the bounding box to include up to 7 km North/South and 14km East/West of the furthest point in the boundary of the municipality on each cardinal direction. The reason for not using the municipality administrative boundaries was two-fold. On the one hand, it allowed me to consider 15 additional meteorological stations which fell outside the administrative boundaries (Figure 4.1), providing additional training points for the model. On the other hand, these training points were of special interest as they captured peri-urban and higher altitude areas, which

were showed a combination of predictor values under-represented in the sample otherwise. Overall, this permitted the model to increase its learning range. We limited the North/South extension (7km) to half of the East/West (14km) because the values of the predictors differed substantially with the rest of the area (highly rural) and there were few or no meteorological station to train and test the model performance on those type of terrains.

Daily temperature was predicted at a 500m^2 regular grid. I decided to use a regular grid over an irregular one for two main reasons. Firstly, the input features, as can be seen below, are all provided in a regular grid format of different resolutions. By predicting also at a regular grid, I aimed to minimize the spatial transformations required to harmonise all input data, lessening the *modifiable area unit problem* [1]. Secondly, a regular grid of such fine resolution (500m^2) offers a great versatility when it comes to its uses as it can be aggregated to other areal units, including irregular areal units such as administrative boundaries for health analyses. This was an important feature for me, as I wanted to ensure the broad use of this dataset by other researchers and purposes further than my thesis.



(a) Study area (b) Meteorological stations

Figure 4.1: **Study area and location of meteorological stations.** (a) Area of study with bounding boxes indicating the contiguous area (orange) and the expanded prediction area (gray). Map of the meteorological stations included (dots) and excluded (triangles) in the model. In grey, the administrative boundaries of the municipality of São Paulo

[1] The modifiable area unit problem is a class of aggregation fallacy by which different ways of drawing a geographic boundary result in vastly different patterns in data measured at that boundary level.

4.2.3 Data sources

4.2.3.1 Temperature data

Supervised machine learning algorithms exploit the associations between different parameters, also known as *input features*, and the target variable to make predictions at unsampled points. To do so, the algorithm first needs to learn these associations using data points for which we have reliable and available estimates of the target variable, what I refer to as the *ground truth*. To maximize the chances of the model to capture the relationships that exists between the input features and the ground truth it is important to have a dense and widespread network of points to use for training. I collected temperature data from eight monitoring networks, which included a total of 60 monitoring stations (Figure 4.1). Temperature is a rather common parameter to measure as it can influence other variables, such as air pollution, and so, it is measured in air pollution monitoring stations as well as meteorological stations. By also including measurements from air pollution monitoring stations, we were able to expand the training points by 8 stations (CETESB network; orange points in Figure 4.1).

4.2.3.2 Input features

In this section, I provide an overview of the predictor data that were considered to build the temperature model. I have classified them thematically in: greenspace, blue-space, atmospheric variables, topographic variables, reanalysis data, temporal variables, and proxies for urbanization. For each group, I list the different variables used, provide a brief technical description of the feature input, and information on the data sources. An overview of all predictor variables used can be found in Table 4.1.

Topography

- ***Digital elevation map***. The areas surrounding the northern part of the municipality of São Paulo are mountainous with some mountain peaks above 1,000m. As our model must be able to predict ambient temperature across the entire spatial extent of the municipality, it is important to include this variable in our model. Here, I used the "NASA SRTM Digital Elevation" version-3 product [255]. The NASA Shuttle Radar Topography Mission (SRTM), see [256] digital elevation data is an international research effort to obtained digital elevation models on a near-global scale. This product is provided at a native resolution of one arc-second ($30m^2$) and three arc-second ($90m^2$). Coarser resolutions are also available, namely $250m^2$ and $500m^2$. I used the $500m^2$ resolution data product to match the target resolution of my model. More information about the data product here. *Specifications: temporal resolution: annual; temporal coverage: 2000-2020; Spatial resolution: 500m^2, and units: meters (m).*

- ***Slope***. The slope represents the rate of change in elevation or inclination of a given terrain and it is calculated in degrees. The values range from 0° to 90°. Here, I derived the slope

data from the DEM product described above, using the GEE function *ee.Terrain.slope*, which calculates the rate of change in elevation using the four connected neighbours of each pixel. *Specifications: temporal resolution: annual; temporal coverage: 2000-2020; Spatial resolution: 500m^2, and units: degrees(°).*

Greenspace features:

- ***Normalized difference vegetation index (NDVI)***. The Normalized Difference Vegetation Index (NDVI) is a scoring system used to classify imagery pixels based on the proportion of vegetation they contain. It is based on the physical property of vegetation to reflect different wavelengths of light. Particularly, the pigment in plant leaves absorb visible light (VIS: 0.4 to 0.7μm) to use in photosynthesis, while near-infrared light (NIR: 0.7 to 1.1μm) is reflected by leaves. The NDVI is the difference in proportion between the VIS reflected in the red part of the visible spectrum and the NIR refracted over the total light exchanged. The greener the landscape, the more VIS light is being absorbed and NIR light refracted and so, the higher the NDVI score. NDVI values range from -1 (lack of green landscape) to $+1$ (green landscape). High NDVI values (0.6-0.9) correspond to dense vegetation such as temperate and tropical evergreen forests or crops at their peak growth. Unhealthy, sparse vegetation or senescing crops have NDVI lower scores (0.2-0.5). I obtained the data from Landsat 7, operative from 1999 to April 6, 2022, a satellite with a sun synchronous orbit of 16-day overpass. The Enhanced Thematic Mapper Plus (EMT+) sensor on board captured eight spectral bands among which, spectral band 4 which captures NIR (wavelength: 0.77 - 0.90μm) and band 3 which captures the reflectance in the red part of the visible spectrum (wavelength: 0.63 - 0.69 μm). Of the available data products, I used 'USGS Landsat 7 Collection 1 Tier 1 calibrated () Reflectance' (daily; 30m) courtesy of the U.S. Geological Survey [255] [2]. I used the 'tier 1' data product as it is the one with the highest available data quality, considered the most suited for time-series analyses. Upon download and data processing in Google Earth Engine (GEE) (see Section 4.2.4.2 below for detailed description of GEE data processing), I then estimated NDVI using the GEE algorithm *ee.Image.normalizedDifference()*, which computes the normalized as:

$$NDVI = (NIR–VIS)/(NIR + VIS) \tag{4.1}$$

$$NDVI = (B4–B3)/(B4 + B3) \tag{4.2}$$

Unfortunately, a large proportion of pixels were masked due to cloud cover. As explained above, NDVI changes over time capturing primarily seasonal variations. Sporadic changes can occur

[2]From December 2022, Landsat 7 Collection 1 datasets have been deprecated. Collection 2 is now the default, offering substantial improvements in the geo-location accuracy.

with sudden alterations of the landscape (e.g., wildfires, harvesting...), yet these are rare. Moreover, the study area was primarily urban and non-agricultural and so, it is unlikely to suffer such abrupt changes in the greenspace. This assumption was tested by visually exploring the day-to-day, week-to-week and month-to-month differences in pixel-wide NDVI score. Differences were minor and so, I aggregated NDVI scores every 3 months. *Specifications: temporal resolution: 3-months; temporal coverage: 2015-2019; Spatial resolution: 500m^2, and units: NDVI score.*

Atmospheric features:

- ***Black sky albedo (bsa)***. When the sun rays enter the earth's atmosphere they are referred to as down welling radiative flux, which has both a direct and a diffuse component. Upon receiving the incident radiative fluxes, the surface of the earth reflects them, known as the up-welling radiative flux. The albedo is the net exchange of radiative flux, or in other words, the ratio of up to down-welling radiative flux at the surface. In particular, Black Sky Albedo (bsa) or directional hemispherical reflectance, refers to the ratio with only the directional component of the down welling flux, as opposed to white-sky albedo or bi-hemispherical reflectance, which includes only the diffuse component of the down welling flux. It is measured on a scale from 0, corresponding to surfaces that absorb all incident radiation, to 1, corresponding to surfaces that reflect all incident radiation. The albedo is affected by local and regional changes and processes, both natural and anthropogenic, including deforestation, wildfires, changes in the snow and ice cover, and presence of aerosols in the atmosphere. Higher bsa values are found in highly populated and industrialized areas; whereas lower values are likely to occur over water bodies and greenspaces. Changes in surface albedo have impacts on both regional and global climate patterns and so, the albedo's spatial and temporal variation is closely related to climate change patterns and regional weather systems. Hence, the albedo is a fundamental parameter for climate modelling since it describes the energy flux at the land boundary layer [257]. Here, I used the 'Moderate Resolution Imaging Spectro-radiometer (MODIS)/Terra+Aqua Bidirectional Reflectance Distribution Function (BRDF) and Albedo Version 6.1.' data product (MCD43A3v061) [258]. It provides both black and white-sky albedo data for every day at local solar noon at 500m^2 spatial resolution. To produce it, it uses 16-days of Terra and Aqua MODIS data temporally weighted to the ninth day of the 16-day window. Data are provided for MODIS bands 1 through 7, as well as visible, NIR, and shortwave bands. Here, I used the shortwave band. The version 6.1 is not available on GEE. Instead, I used the NASA's Application for Extracting and Exploring Analysis Ready Samples (AppEEARS) platform Python API [259]. The data downloading, processing and exporting was similar to that described for GEE, yet using AppEEARS specific algorithms and functions. *Specifications: temporal resolution: daily; temporal coverage: 2015-2019; Spatial resolution: 500m^2, and units: standardized score, range: 0-1.*

- ***Solar Zenith Angle (SZA)***. When observing a given point on the Earth's surface from a satellite-based instrument, the Solar Zenith Angle (sza) is the angle between the local zenith (i.e. directly above the point on the ground) and the line of sight from that point to the sun. This means that the higher the Sun is in the sky, to lower the sza. The angle between the local zenith and the line of sight to the satellite, is called the Viewing Zenith Angle. It has been shown to be sometimes a good predictor for temperature modelling [111, 260, 261]. I used the 16-day sun zenith angle band of the "MODIS/Terra Vegetation Indices 16-Day L3 Global 1km SIN Grid" USGS data product [262]. Imagery was processed as indicated in Section 4.2.4.2, using GEE's Python API. *Specifications: temporal resolution: daily; temporal coverage: 2015-2019; Spatial resolution: 1km^2, and units: degrees (°).*

- ***Land surface temperature (LST)***. Land Surface Temperature (LST) or radiative skin temperature of the land refers to the thermal radiance emitted by the Earth's surface as a result of its heating from the incoming solar radiation. It plays a key role in surface energy balance and evapotranspiration processes, amongst others. LST can be measured *in situ*, modelled or retrieved from remote sensing. The latter is by far the most used one as it provides global and continuous coverage, long temporal series and high spatio-temporal resolution at a relatively low expense. There are a wide variety of LST products available, each with different temporal and spatial resolutions. Landsat LTM, ETM+, and ASTER are the satellites currently providing the highest spatial resolution, between 30m-120m, yet this is at the expense of a coarser temporal resolution, between 2 and 16 days. These are often used for UHI studies, where the spatial component is more important than the temporal component. The MODIS [263], Visible Infrared Imaging Radiometer Suite (VIIRS) [264], and Sea and Land Surface Temperature Radiometer (SLSTR) LST [265] provide a coarser spatial resolution of 1km^2 but at a much finer temporal resolution; daily. This is especially of interest for time-series analyses and epidemiological studies, where the temporal component is a vital component. Among them, MODIS LST is the most widely used for its long time series (available since 2000), and its well-calibrated and documented data records [266–268]. Here, I used the process described by Shiff et al.,'s 2021 [269] approach to impute missing values. Refer to Section 4.2.4.2 for more detail on the methods for the spatio-temporal data imputation. *Specifications: temporal resolution: daily; temporal coverage: 2015-2019; Spatial resolution: 500m^2, and units: degrees Celsius (°C).*

Blue-space features:

Water has a cooling effect on air temperature [270]. The magnitude of the cooling effect is related to the size of the water body, larger water bodies have bigger impacts; the land cover of the surrounding areas, with built-up areas having shorter catchment areas and effect sizes; the water/air temperature ratio, and the incident solar radiation, wind speed and relative humidity[270]. The effect of small water

bodies in intra-urban areas remains controversial regarding the importance of its cooling effect[270]. Therefore, here I only focused on large water bodies such as lakes or dams, or the ocean, ignoring smaller bodies such as rivers, canals or fountains.

- ***Distance to inland water.*** The municipality of São Paulo falls within the Tietê River Basin in the Paraná Hydrographic Region. There are three major dams intersecting the municipality, the *Represa de Guarapiranga*, the *Represa Rio Pedras,* and the *Represa Billings* (Figure 4.2 (a) and (b). I downloaded the vector layer 'Represa-Nivel Maximo" from the Alto Tietê data platform, managed by the *Comitê da Bacia Hidrográfica do Alto Tietê* [271]. This layer captures the maximum capacity of the two aforementioned dams. For each dam, I drew a 200 and 400m distance buffer (Figure 4.2(b)). I then overlaid the buffers to a regular target grid of 500m^2 For each grid cell, I estimated the percentage of area intersecting the buffer and assigned it a score, with 100 representing a cell fully intersecting the buffer and 0, not at all. The result were three independent dummy rasters containing information on the proportion of each grid cell intersecting the dams at 200m, and at 400m of distance (Figure 4.2(c), (d) and (e), respectively). *Specifications: temporal resolution: static; temporal coverage: 2000-2020; Spatial resolution: 500m^2, and units: meters (m).*

- ***Distance to the coastline***. I estimated the distance to the coastline using R software. Firstly, I imported Brazil's coastline profile from the *rnaturalearth* package which imports the boundaries of countries from around the world from the platform Natural Earth [272]. The data original coordinate system is WGS84 with decimal degrees (EPSG code: 4326). For the calculation of distances, it is more convenient to use meters instead of degrees. Thus, I transformed the coastline layer to the regional UTM Zone 23 (EPSG: 31983). We then need to define the points where we want to know the distance to the coastline for. In my case, I used the centroids of my target raster with 500m^2 grid cells and spatial extent of lon_{min}: -46.9559; lat_{min}: -24.0854; lon_{max}: -46.2226; lat_{max}: -23.2839, coordinate system: WGS84. I calculated the centroid-to-coastline distance using the *st_distance()* function, which outputs a vector of distances for all our points. This is then transformed to a data frame containing the geo-referenced location of each point and the estimated distance. Finally, this data frame is transformed back to a raster using the *rasterize()* function. *Specifications: temporal resolution: static; temporal coverage: 2000-2020; Spatial resolution: 500m^2, and units: meters (m).*

Reanalyses data: ERA5-Land

Reanalysis refers to the scientific method used for developing comprehensive historical records of weather and climate variables by combining numerical models with climate observations. The same climate model is used through the entire reanalysis period to reduce variability due to modelling and

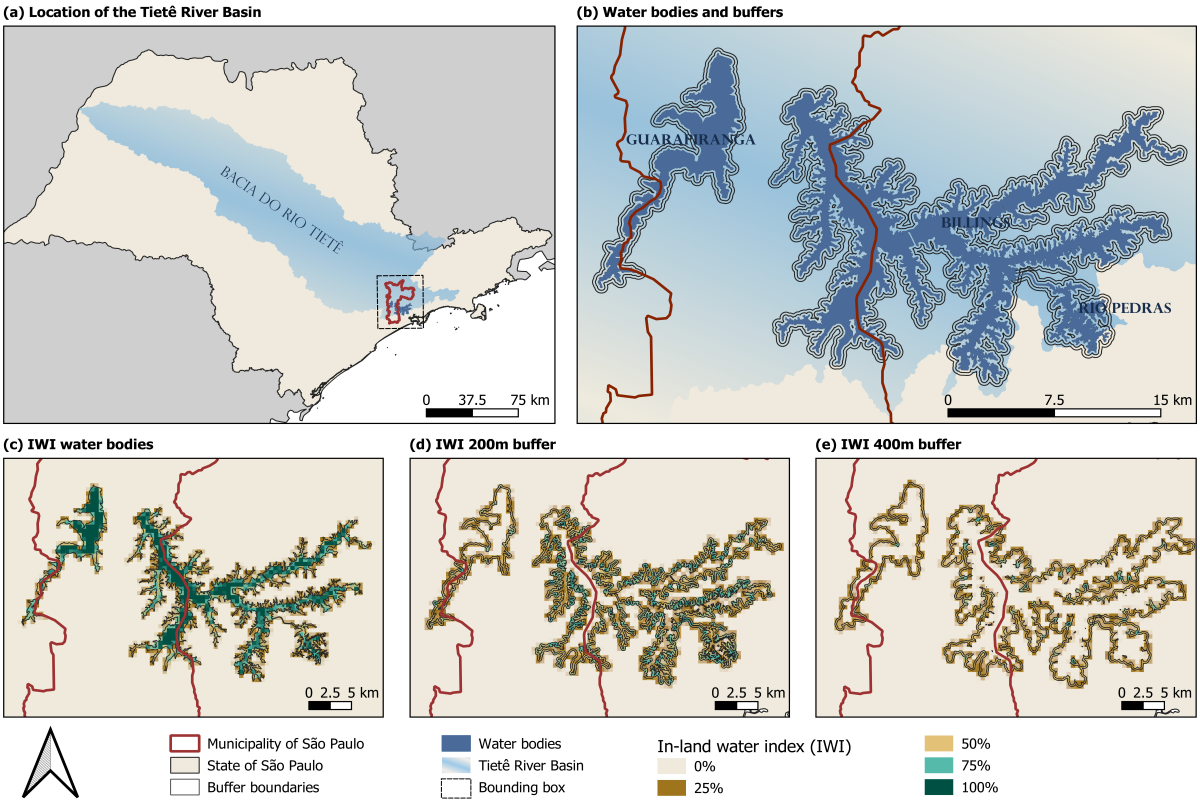


Figure 4.2: **River index variable.** (a) In light blue, the *Tiete* river basin expanding over the State of São Paulo. (b) Close up view of the three major inland water bodies; the *Guarapiranga*, the *Billings* and *Rio Pedras* dams in dark blue. The 200m and 400m distance buffers are shown in black. (Bottom) Maps showing the percentage of area per grid cell intersecting (c) the water bodies, (d) the 200m buffer, and (e) the 400m buffer.

statistical choices. The ERA5-Land reanalysis dataset[117] provides hourly information land variables only, for years between 1950 and the present, at 9km^2. The data is a replay of the land component of the ERA5 climate reanalysis. ERA5-Land does not use observations directly, instead it uses ERA5 atmospheric variables, e.g., air temperature and humidity, to define the atmospheric forcing constrain. More information on the methods to generate and calibrate ERA5-Land can be found in Munoz-Sabater et al., 2021 [117]. Again, ERA5-Land also includes information about uncertainties for all variables. The selection of the ERA5-Land variables was based on their potential to explain the spatial and/or the temporal variability of ambient temperature in the municipality of São Paulo. Below, I describe them in detail. *Specifications: Temporal resolution: daily; temporal coverage; 1950-present; spatial resolution: 9km^2. These apply to all variables and are stated here to avoid repetition.*

- ***2m temperature (t2m).*** This variable represents the ambient temperature at a height of two meters above the Earth's surface. To calculate it, they interpolate the Earth's surface temperature and the lowest model value, considering atmospheric conditions. *Units: Kelvin(K).*

- ***2m dew point temperature (d2m).*** This variable is the temperature at which the air would

have to be cooled for saturation to occur at two meters above the Earth's surface. Thus, it measures the humidity of the air. *Units: Kelvin (K).*

- ***Skin temperature (skt)***. This variable is the theoretical temperature required to satisfy the surface energy balance. It captures the temperature at the top surface layer, which has no heat capacity and so responds almost instantaneously to changes in surface fluxes. *Units:Kelvin (K).*

- ***Relative humidity (rh).*** This variable is the relative humidity at two meters above the Earth's surface. It was calculated from the 2-meter air temperature (t2m) and the 2-meter dew point temperature (d2m) using the formula for saturated water pressure from Wright (1997) specified below, where t2m and d2m are measured in degrees Celsius (°C). *Units: percentage (%).*

$$\text{Max. water pressure} = 611.21 * \exp[17.502 * t2m/(240.97 + t2m)] \tag{4.3}$$

$$\text{Actual water pressure} = 611.21 * \exp[17.502 * d2m/(240.97 + d2m)] \tag{4.4}$$

$$\text{RH} = \text{actual water pressure/maximum water pressure} \tag{4.5}$$

- ***Surface pressure (sp).*** This variable is the pressure of the atmosphere at an area of the Earth's surface. It is measured as the weight of air in a vertical column in force per unit area. *Units: Pascals (Pa).*

- ***10m V-wind component (v10).*** This variable captures the horizontal speed of air moving north at a height of ten metres above the Earth's surface. *Units: meters per second (m/s).*

- ***10m U-wind component (u10).*** This variable captures the horizontal speed of air moving east at a height of ten metres above the Earth's surface. *Units: meters per second (m/s).*

Temporal variables:

- ***Day length***. We define day length as the number of hours in a day with direct and indirect sunlight. Sunlight is positively correlated often with warmer temperatures, i.e. daytime temperatures are usually higher than night-time temperatures. Since we are modelling 24h mean temperature, the average temperature will be affected by the proportion of hours with sunlight. Finally, the number of hours of sunlight is a proxy for seasonality, another key explanatory variable for ambient temperature. *Specifications: temporal resolution: annual; temporal coverage: 2000-2020; Spatial resolution: municipality of São Paulo, and units: hours (h).*

Proxies for urbanization:

- ***Population density***. Usually population density is obtained from census data, and made available as counts per administrative unit. These irregular units may mask small variations

and can cause a spatial mismatch problem when integrated with other datasets. To avoid these problems, I used the gridded global population density dataset provided by the WorldPop platform [273]. WorldPop is a research program set up in 2013 at the School of Geography and Environmental Sciences at the University of Southampton. Its multi-disciplinary team produces open-source data on global population distribution at high spatial resolution. They use a top-down approach to bring administrative unit-based population count data from census and their projections to grid cell-based counts. To do so, they developed two Random Forests machine learning approaches, the unconstrained and the constrained approaches, described in Stevens et al., 2015 [274]. The constrained approach assumes that we have good quality data of the location of human settlements, including small ones and so, that these can be used to delimit the spatial distribution of the population density data layer, masking areas where we do not expect settlements. The unconstrained approached, on the contrary, assumes that no settlement dataset is accurate enough to identify all residential settlements and therefore, it cannot be used to mask the population density data layer. Once population counts have been downscaled either using the unconstrained or the constrained approach, they can be used to estimate the population density per pixel by dividing the number of people in each pixel by the pixel surface area. Informal settlements are common in the municipality of São Paulo, also known as *favelas*. These areas are of irregular size, shape and infrastructure, which makes them more likely to go amiss in settlement datasets. People living in these settlements are often of low socioeconomic status, live in poorer conditions and are more vulnerable to suffer from poorer health, making them a particularly relevant group to capture. Despite recent refinements are now making feasible the detection of informal settlements, to avoid missing important areas with particularly vulnerable population, I decided to use the unconstrained data product. *Specifications: temporal resolution: annual; temporal coverage: 2000-2020; Spatial resolution: 1km^2, and units: number of people per square kilometre (people/km^2).*

- ***Impervious surfaces***. A surface is impervious when it is constructed with water impenetrable materials such as asphalt, concrete, brick or stone. Densely compacted soil, such that found in urban plots, can sometimes also behave like an impervious surface for its water-resistant quality. In addition to sealing the soil surface, hindering water infiltration processes, impervious surfaces have the property of storing thermal radiation and releasing it during the night time. This phenomenon gives raise to urban heat islands. Its role in temperature modulation is clear and thus, it was explored as an input feature. Impervious surfaces are a direct consequence of urban sprawl and so, their abundance and spatial distribution is likely to vary over time with increasing urbanization. Consequently, impervious surfaces and population density are highly correlated.

  The population density dataset used has an annual temporal resolution and a 1km^2 spatial res-

olution, capturing the areas urban sprawl over time but failing to detect small settlements or complex spatial patterns. Thus, I deemed appropriate the use of a impervious surface dataset at much finer spatial resolution to compliment the population density data, at the expense of a lower temporal resolution. The dataset used was developed by Zhang et al (2020) [275] and it is publicly available globally at a $30m^2$ resolution for year 2015. The authors build a Random Forest Classifier to generate $30m^2$ resolution maps with a binary classification (1, impervious surfaces). As input features, they used data from three typologies of satellite sensors: Landsat 8 Operational Land Image (OLI) optical images, which provide information on surface reflectance; Sentinel1 synthetic-aperture radar (SAR) images, which provide information on structure and dielectric properties of the materials, and Visible Infrared Imaging Radiometer Suite (VIIRS) nighttime light images, which detect human activity, highly suggestive of the presence of impervious surfaces. As ground truth to train the models, they used a combination of three widely used and validated datasets: GlobeLand30, VIIRS NTL and MODIS EVI data. They achieved an accuracy of 95.1%, outperforming other impervious surface datasets and methods available to date. For more details on the methodology used, refer to the paper [275]. *Specifications: temporal resolution: 2015; temporal coverage: 2015; Spatial resolution: 30m^2, and units: binary (1, impervious surface; 0, non-impervious surface.)*

-

Table 4.1: **Candidate input features.** List of input features considered for inclusion in the Random Forest model with information the data product names, original and post-processing temporal and spatial resolution, percentage of missing values (NAs), imputation method used if any, temporal coverage of the data product, and data source reference. S: spatial, ST: spatiotemporal and T: temporal

	Overall group	Variable	*Acronym*	Product name	NAs(%)	Imputation method	RMSE	Original Spatial Resolution	Original Temporal Resolution	Temporal coverage	Ref
S	Anthropogenic heat proxies	Impervious surfaces	*impsurf*	Dataset by Xiao Zhang et al., 2020	0 %	-	-	$30m^2$	Annual (2015)	2015	[275]
ST	Anthropogenic heat proxies	Population density - WorldPop	*popdens*	WorldPop	0%	-	-	$1km^2$	Annual (2015)	2015	[273]
S	Terrain	Elevation	*dem*	SRTMGL1 v003	0%	-	-	$500m^2$	Annual (2015)	2015	[255]
S	Terrain	Slope (Inclination)	*slope*	SRTMGL1 v003	0%	-	-	$500m^2$	Annual (2015)	2015	[255]
S	Terrain	Land cover	*landcov*	ESACCI-LC: European Space Agency Climate Change Initiative Land Cover product	0%	-	-	300^2	Annual (2015)	2015	[276]
S	Green and blue spaces	Distance to coastline	*coast*	Estimated	0%	-	-	.shp coast	Annual (2004)	2004	[277]
S	Green and blue spaces	Distance to water bodies	*water*	Estimated	0%	-	-	.shp water	Period (2000-12)	Period (2000-2012)	[272]
S	Green and blue spaces	Distance to lakes (lake boundary)	*lake*	Estimated	0%	-	-	.shp dams	Annual (2004)	2004	[271]
S	Green and blue spaces	Distance to lakes (200m buffer)	*b200m*	Estimated	0%	-	-	.shp dams	Annual (2004)	2004	[271]
S	Green and blue spaces	Distance to lakes (400m buffer)	*b400m*	Estimated	0%	-	-	.shp dams	Annual (2004)	2004	[271]
ST	Green and blue spaces	Normalized difference vegetation index (NDVI)	*ndvi*	USGS Landsat 7 Collection 1 Tier 1 Reflectance	5.05%	Gapfill imputation (Gerber et al., 2018)	0.05	$500m^2$	Daily	2015-2019	[278]
ST	Anthropogenic heat proxies	Distance to artificial land cover	*artland*	WorldPop	0%	-	-	$100m^2$	Annual (2015)	2015	[273]
ST	Anthropogenic heat proxies	Population density - WorldPop	*popdens*	WorldPop	0%	-	-	$1km^2$	Annual (2015)	2015	[273]
ST	Atmospheric variables	Black Sky Albedo	*bsa*	$MCD43A3_v006$	13.0%	Gapfill imputation (Gerber et al., 2018)	0.01	$1km^2$	Daily (based on 16-day average)	2015-2019	[258]
ST	Atmospheric variables	Land surface Temperature	*lst*	$MOD11A1_v061 + CFSv2$	-	LST dataset by Shiff et al., 2021[269]	2.38	$1km^2$	3-times per day	2015-2019	[269]
ST	Atmospheric variables	Solar Zenith Angle	*sza*	$MOD13A2_v006$	4.67%	Gapfill imputation (Gerber et al., 2018)	1.94	$1km^2$	16-day average	2015-2019	[262]
ST	Weather	Temperature at 2m; ERA5-Land	*t2m*	ERA-5 Land	0%	-	-	0.1° x 0.1°; Native resolution is 9 km	Daily	2015-2019	[115]
ST	Weather	Wind speed; ERA5-Land	*v10*	ERA-5 Land	0%	-	-	0.1° x 0.1°; Native resolution is 9 km	Daily	2015-2019	[115]
ST	Weather	Wind direction; ERA5-Land	*u10*	ERA-5 Land	0%	-	-	0.1° x 0.1°; Native resolution is 9 km	Daily	2015-2019	[115]
ST	Weather	Pressure; ERA5-Land	*sp*	ERA-5 Land	0%	-	-	0.1° x 0.1°; Native resolution is 9 km	Daily	2015-2019	[115]
ST	Weather	Skin temperature; ERA5-Land	*skt*	ERA-5 Land	0%	-	-	0.1° x 0.1°; Native resolution is 9 km	Daily	2015-2019	[115]
ST	Weather	Relative humidity; derived from ERA5-Land	*rh*	ERA-5 Land	0%	-	-	0.1° x 0.1°; Native resolution is 9 km	Daily	2015-2019	[115]
ST	Weather	Dew point 2m temperature; ERA5-Land	*d2m*	ERA-5 Land	0%	-	-	0.1° x 0.1°; Native resolution is 9 km	Daily	2015-2019	[115]
T	Temporal variables	Day length	*daylength*	Estimated	0%	-	-	-	Daily	2015-2019	-

4.2.4 Pre-modelling data wrangling

4.2.4.1 Weather stations data processing

Operational period: Initially, I collected data from 60 meteorological stations, each with different operational periods. As an initial assessment, I plotted their data completeness by year (Figure 4.3). Based on the availability of data, I restricted the study period between 2015 and 2019, for it had the highest proportion of available stations. Hereafter, all data refers to this period and stations.

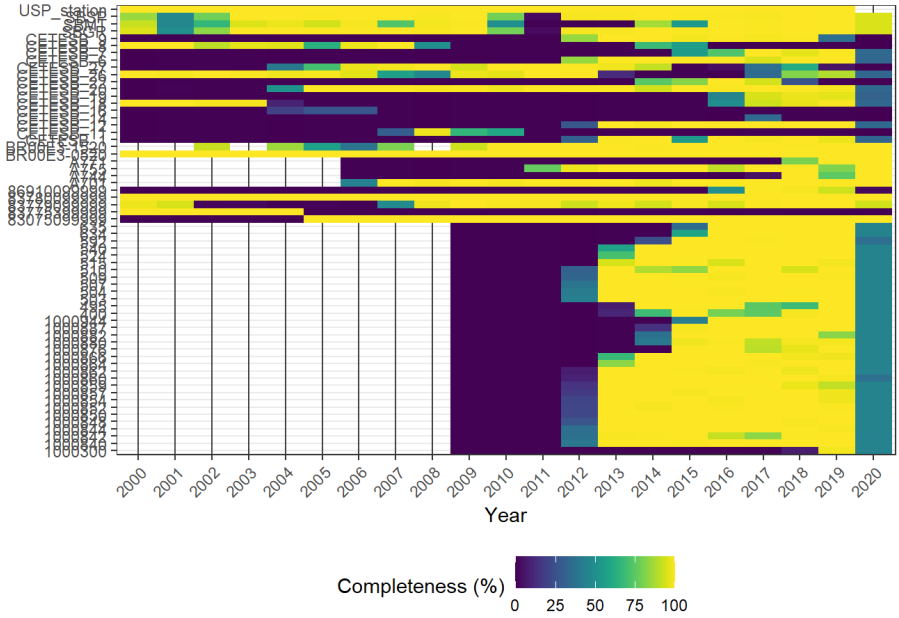


Figure 4.3: **Heatmap of annual data completeness by station.** Heatmap of annual completeness as a percentage (%) of the total annual observations by station between 2000 and 2020.

Spatial clustering: Prior to the data cleaning, I clustered all stations based on their spatial proximity using k-means spatial clustering. K-means clustering [279] is one of the most used unsupervised machine learning algorithms for partitioning a given data set into a set of k groups, where k represents the number of groups pre-specified by the analyst. It classifies objects in multiple groups, referred to as clusters, such that objects within the same cluster are as similar as possible (i.e., high intra-class similarity), whereas objects from different clusters are as dissimilar as possible (i.e., low inter-class similarity). In other words, each observation (x_i) is assigned to a given cluster such that the sum of squares (SS) distance of the observation to their assigned cluster centres μk is a minimum. In k-means clustering, each cluster is represented by its centre or centroid, which corresponds to the mean of points assigned to the cluster. In my case, the feature used to cluster the stations was the pairwise physical distance between the monitoring stations.

To identify the optimal number of clusters, I iteratively ran the algorithm using different values of k. The weighted sum of squares for each iteration is then plotted to identify which k value produces the minimum sum of squares. Elbow plots are commonly used for this purpose (Figure 4.4 (a)). It

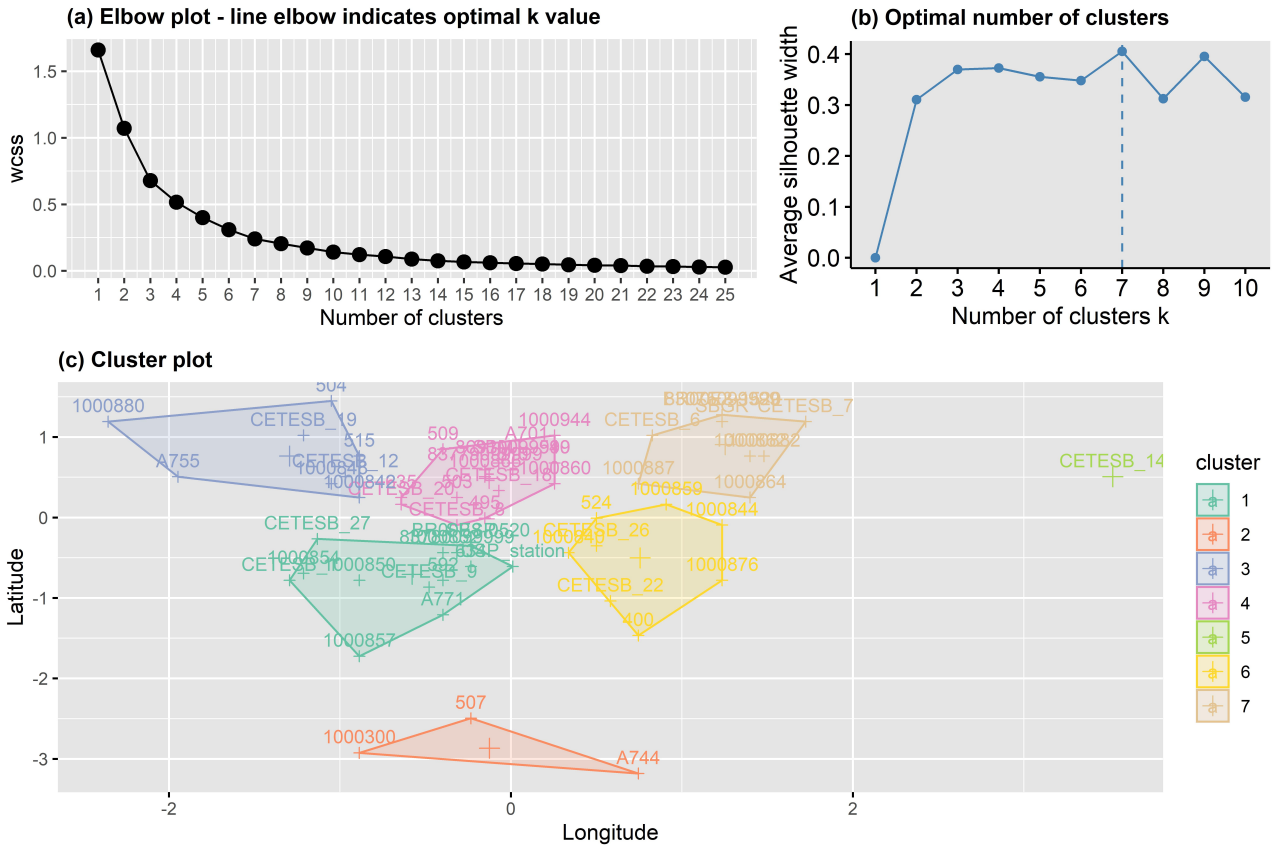


Figure 4.4: **Meteorological station spatial clusters.** Elbow plot showing the weighted sum of squares for k number of clusters (top left); Line plot of the average silhouette width by k number of clusters (top right); Optimal number of clusters (k=7) mapped and coloured by cluster (bottom).

is a line plot with each of the iteration's Within-Cluster-Sum of Squared Errors (WSS) plotted. The point where the line starts to flatten, the so-called inflection point, indicates the optimal k value. This can be hard to visualize sometimes. In those cases, the Silhouette plot can be used(Figure 4.4 (b)). It assigns a value from 1 to -1 to each iteration based on the similarity within clusters and across clusters; the higher the value the higher the similarity within clusters and so the more optimal the k value is. Below, I present the results from testing 2:25 k values. I selected k=7 as the optimal number of clusters and used that value to cluster all the stations. Figure 4.4 (c) shows the spatial clusters derived from this procedure.

Duplicated stations: Stations were considered duplicate if their latitude and longitude coordinates were the same to the third decimal place. Upon identifying a potential suspect of duplicated stations, I confirmed the location in the data providers documentation. There were two pairs of duplicate stations (BR00E3-0520 and SBSP; BR00E3-1520 and 830750999999). I retained the station with a more complete series. For the first pair (Figure 4.5 (a)), despite showing similar oscillations, the BR00E3-1520 station had remarkably higher temperatures (+10 °C). The differences were large enough to suspect that there was a data issue with one of the stations. To identify which station was the one affected, we compared them to their nearby stations using the spatial clusters defined in the

section above. We would expect the temperatures to be similar with stations that are spatially close. The BR00E3-1520 consistently reported temperatures 10 °C above the other stations in the cluster, and so, it was excluded. For the second pair (Figure 4.5 (b)), the GHCN station (BR00E3-0520) was more complete and thus, retained.

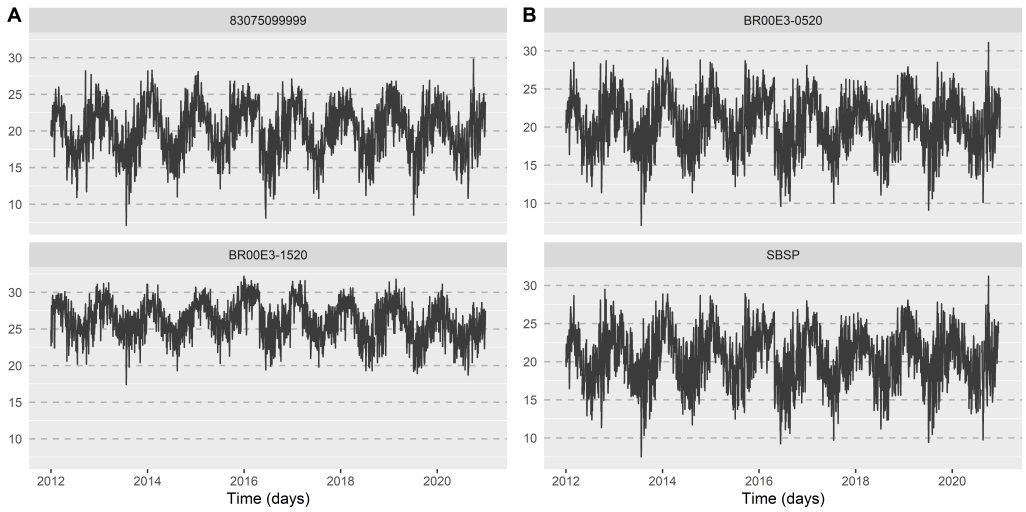
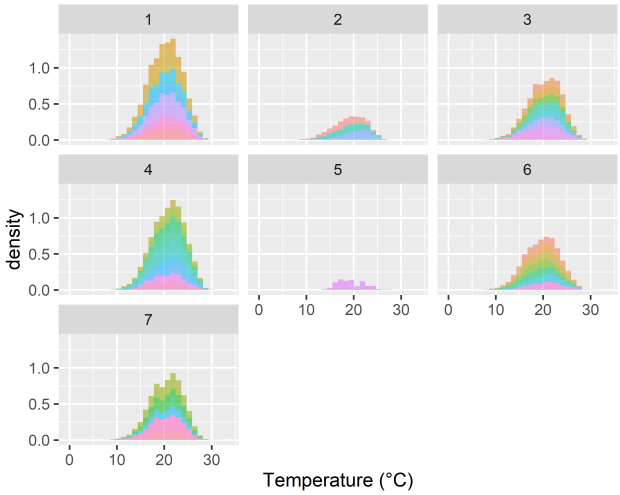


Figure 4.5: **Meteorological stations duplicates.** Time series of the duplicate stations BR00E3-0520 and SBSP (A) and BR00E3-1520 and 830750999999 (B).

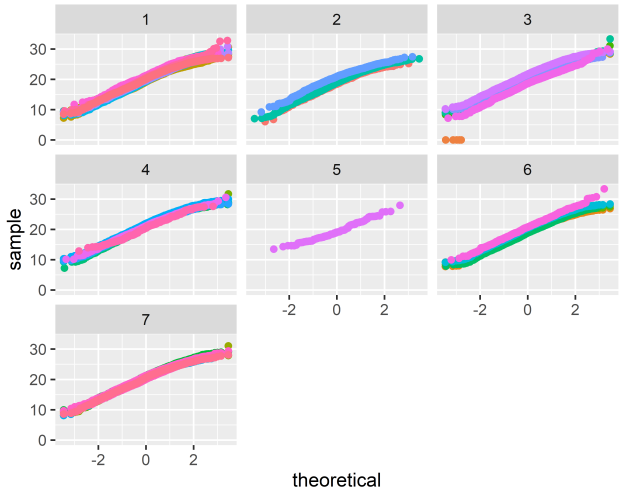
Substitution bias: Several networks report applying basic data imputation methods to fill in gaps in the data. One such method is to fill in missing values with the value recorded on the nearby day. If the window of missing values is large, this can lead to important biases. We refer to this as *substation bias* hereafter. I considered an observation to be affected by substitution bias if the same exact temperature estimate is recorded for more than two consecutive days and for which a similar pattern was not observed in any other station of the same cluster. This can be visually examined using the time series and correlation plots. When such cases were encountered, the affected values were set to *NULL* except for the first one in the series, assuming that it was the true value used to impute the missing following days. I identified two stations affected by at least one event of substation bias.

Outliers: Generally, an outlier is an observation that lies an abnormal distance from other values. However, there is no consensus on the exact definition of outliers as it varies depending on the variable of interest and its use. For environmental variables for which extreme values have remarkable health consequences, such is the case of temperature, it is important to consider whether such deviations are indeed data collection/quality issues or real extreme events. In our case, I considered an observation to be a true outlier if the value met both criteria: (i) was at least two degrees above or below the maximum or minimum percentile of a given cluster, respectively, and (ii) was the only observation at that position, i.e., no other station recorded an outlier with a similar value (± 1 °C). To identify outliers in the data, I used three complementary tools applied to each individual station, i.e., histograms,

normal Q-Q plots and boxplots (Figure 4.6, and 4.7 respectively).



(a) Histograms by cluster



(b) Normal Q-Q plots by cluster

Figure 4.6: **Histogram and normal Q-Q plot of temperature by cluster.** (a) Histogram and (b) normal Q-Q plots of the daily mean temperature recordings from each meteorological stations (colors) grouped by spatial cluster ($1 \rightarrow 7$).

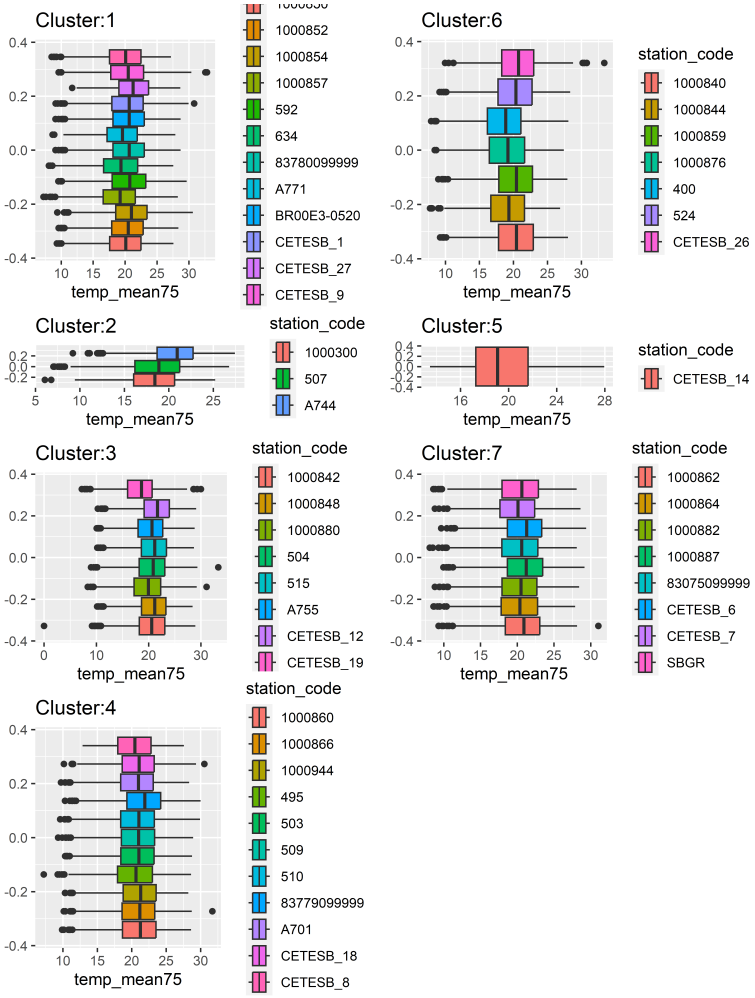


Figure 4.7: **Box plot of station-specific temperature by cluster.** Box plots of the daily mean temperature recordings from each meteorological stations (colors) grouped by spatial cluster ($1 \rightarrow 7$).

Completeness: RF regression algorithms can handle missing values in the ground truth during the training. Nonetheless, the presence of missing data it is undesirable as it reduces density of data points for training and can introduce a bias in the model if the presence of missing values is more prevalent in stations of certain characteristics. And so, it is important to understand not only the proportion but also the distribution of missing values across the time series. According to this distribution, I defined three classes of missing values:

- **Class 1:** Missing values before and/or after the operational period, which undoubtedly hinder the number of training points but do not suggest a data quality issue.
- **Class 2:** Big windows of missing data during operating period, likely to be associated with faulty or broken equipment. Such events may be arbitrary or associated to vandalism.
- **Class 3:** Finally, missing observations or observations classified as erroneous (e.g., biased, outliers) that occur spread out during the operating period, likely associated to equipment or operator errors. A large proportion of the latter is suggestive of a poor-quality station.

When defining the completeness criteria, it was important to find a balance between retaining enough station-specific data and ensuring that the retained information was of high quality and not biased. Firstly, I excluded years for which there was no recording during the study period and stations with less than one year of data (n=10). This addressed the presence of Class 1 missing values. Secondly, I excluded stations which provided less than 75% of valid temperature recordings for the entire study period (2015-2019) (n=2). This restriction tried to ensure that there would be enough training points to capture the associations existing in the unique feature space that was captured by that station. For class 2 and 3, it is important to asses whether missing data are disproportionately affecting a certain set of stations or time periods, as this may introduce a bias to the training dataset. To assess this, I investigated whether they had a differential behaviour over time (day of the week, month, year), network, space (clusters) and area-level characteristics (rurality, deprivation).

- **Temporal bias**. Figures 4.8 and 4.9 show the percentage (%) of missing values for each station by year, month and day of the week, respectively. I observed no consistent pattern in the distribution of missing values in neither of the temporal scales examined, confirming the absence of a temporal bias.
- **Network bias**. Figure 4.10 (a) shows the percentage (%) of missing values for each station by network. It seems like CETESB is the network with a higher percentage of missing temperature recordings. This is primarily an air pollution network which may explain the lack of prioritization of the temperature recordings when a machine is faulty or damaged. The INMET network of stations also showed a large proportion of missing data, yet in this case it was primarily due to

the operational date falling outside the study period for two out of four stations. The rest of networks show a good annual coverage for the operating period.

- **Spatial bias.** Figure 4.10 (b) shows the percentage (%) of missing values by cluster. Cluster 2 and 5 suffer from the largest proportion of missing values, which is in line with the network findings as these two clusters are primarily formed by CETESB and INMET stations.

- **Area-level characteristics bias.** I explored the correlation between percentage of missing values and the SES index (Figure 4.11 (a)) of the area where the monitor was found. Devises can be damaged or broken by vandalism, which is more likely to occur in areas of high socioeconomic deprivation. From Figure 4.11 (a), there did not seem to be any clear difference between the proportion of stations with missing data by SES group. Finally, I investigated the correlation of missing values with the level of urbanicity of the areas where the monitors are found (Figure 4.11 (b)). Proportionally, there were more stations for which the first years were not available as they were not operative yet. This likely reflects the small population density and thus, need for monitoring in rural settings.It is possible that the lack of services and longer distances may be associated to longer windows of missing data as broken or faulty devices take longer to be repaired. From my investigations, there was no obvious difference between rural and urban stations when it comes to the distribution of missing data.

At the end, I had 43 ground monitoring stations with a total of 87,648 valid observations to use as ground truth to train my RF model.

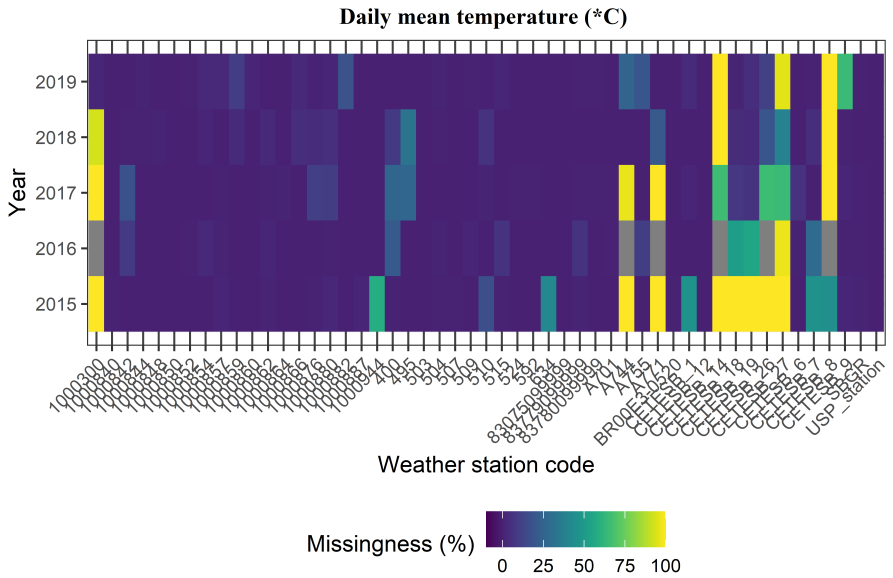


Figure 4.8: **Station-specific missing values by year** Percentage of daily missing temperature values for each station (*x-axis*) and year (*y-axis*).

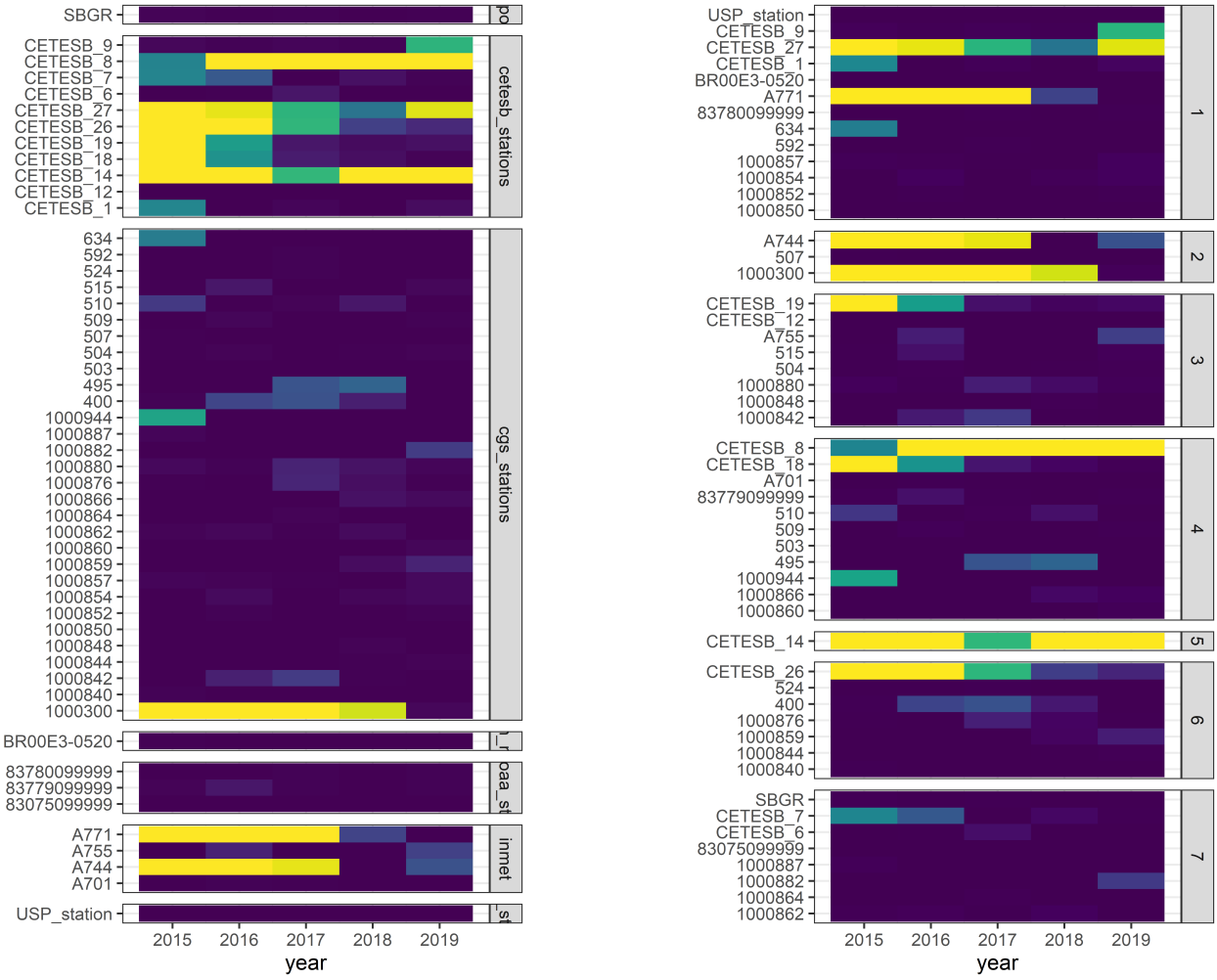


(a) Missing values by month



(b) Missing values by day of the week

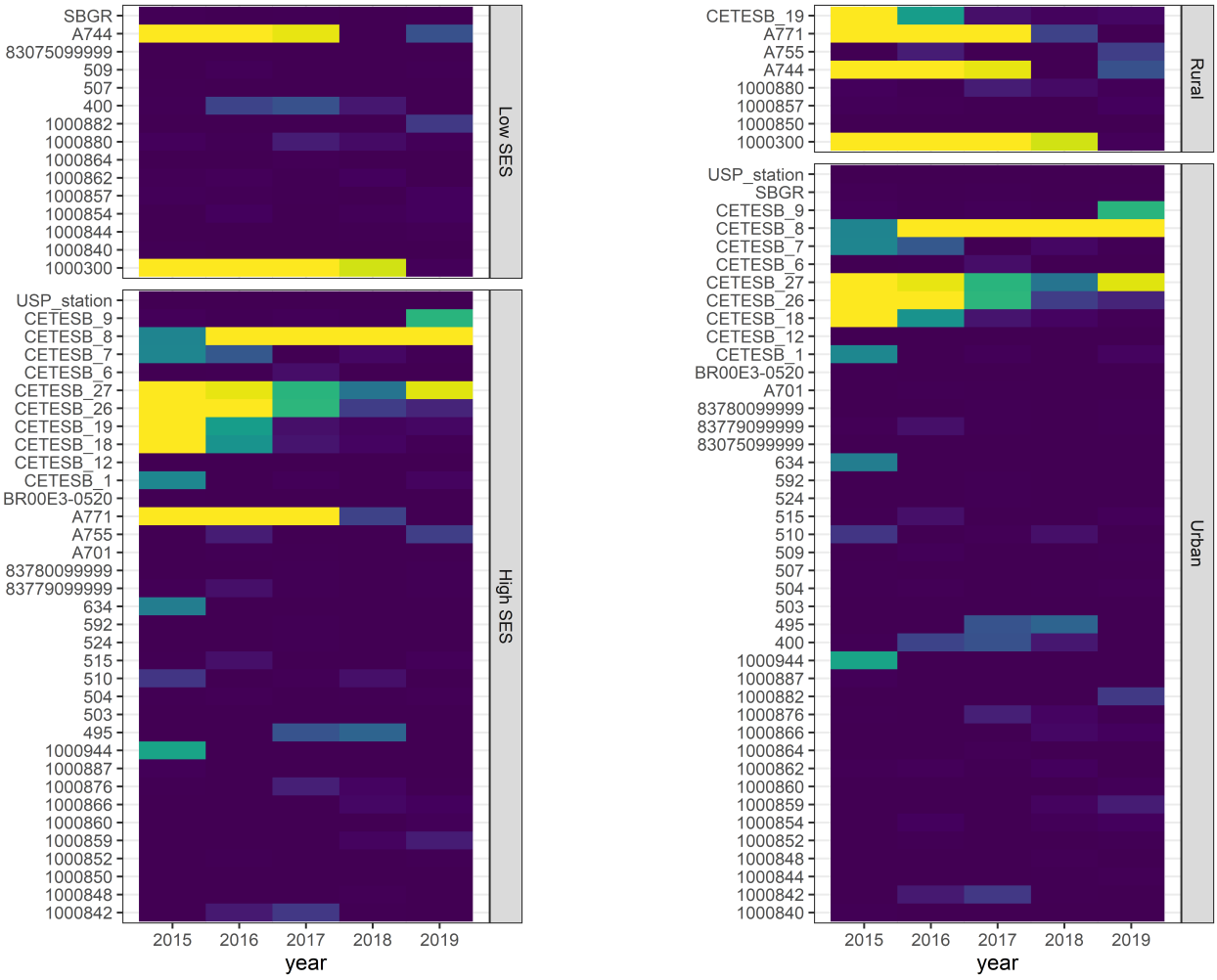
Figure 4.9: **Station-specific missing values by month and day of the week.** Percentage of daily missing temperature values for each station by (a) month and (b) day of the week (*y-axis*) across each year (*x-axis*).



(a) Missing values by network

(b) Missing values by cluster

Figure 4.10: **Missing temperature by network and cluster.** Percentage of daily missing values by year and station, grouped by (a) meteorological station network, and (b) spatial cluster.



(a) Missing values by SES

(b) Missing values by rural/urban

Figure 4.11: **Missing temperature by SES and rurality.** Percentage of daily missing values by year and station, grouped by (a) category of socioeconomic deprivation and (b) rural or urban classification of the station.

4.2.4.2 Input feature data processing

Here, I describe in detail the process followed to select and download the satellite data, the approach taken to identify and impute missing data, and finally, the methods applied to harmonize all data spatially and temporally to match the target resolution (daily; 500m^2). Figure 4.12 represents the workflow followed.

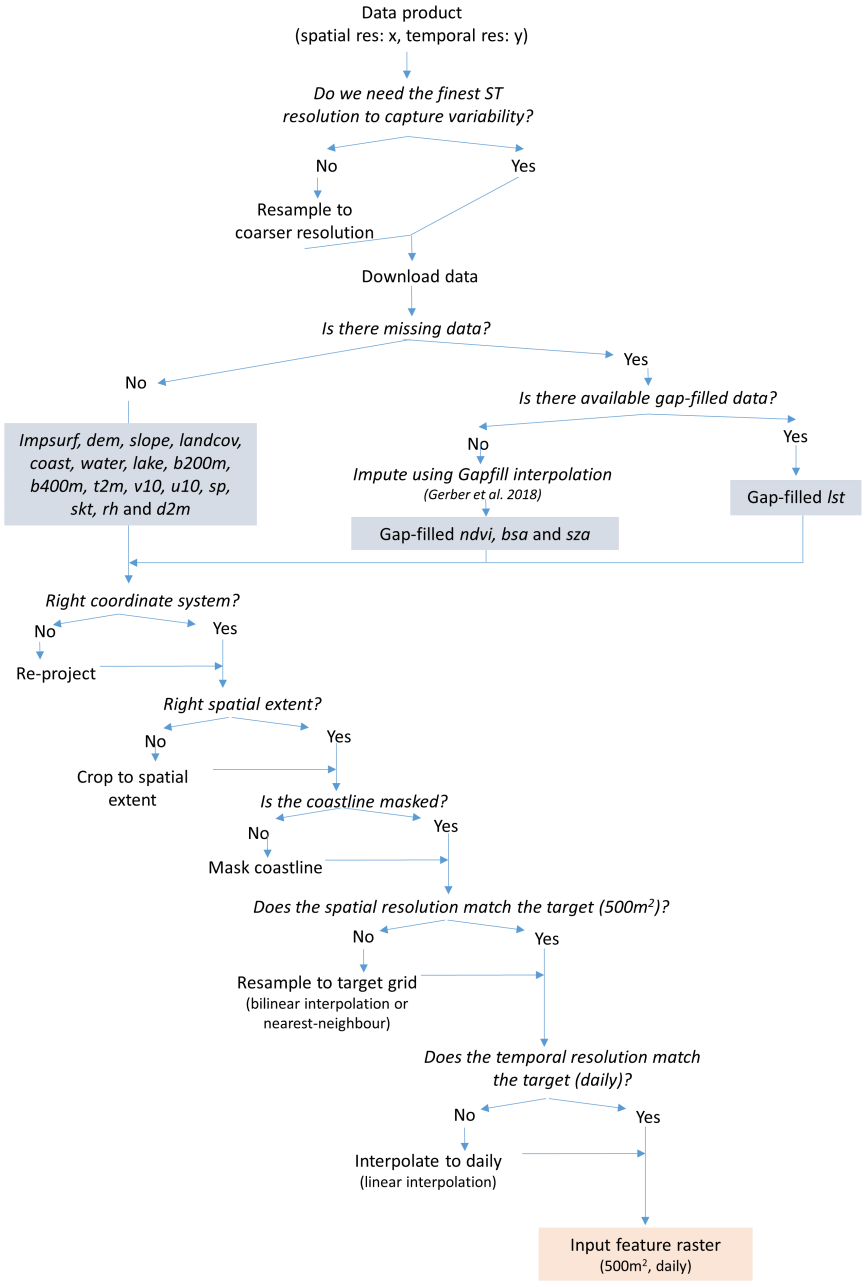


Figure 4.12: **Workflow of the imputation and spatiotemporal harmonization of features.** Visual representation of the workflow followed to assess the need for data imputation and spatial and temporal harmonization of all considered input features.

Satellite and satellite-derived data collection and initial processing

Most satellite and satellite-derived products were downloaded from Google Earth Engine (GEE) platform [280] using the Python API, except for the bsa, for which the latest version was not available in GEE at the time of the download. GEE is an advanced cloud-based geo-spatial processing platform, powered by Google Cloud Platform. GEE provides access to more than 30 years of historical sensor satellite images and ready-to-use satellite-derived datasets, regularly updated and expanded. It also incorporates a wide range of suitable algorithms to process, visualize, analyse, extract, and create spatial data. For each satellite or satellite-derived product, I proceeded as follows:

1. **Define area of study**. I defined the study area using a rectangular bounding box that matched our area of prediction (lon_{min}: -46.9559; lat_{min}: -24.0854; lon_{max}: -46.2226; lat_{max}: -23.2839, coordinate system: WGS84; Figure 4.1).

2. **Define dates of interest**. I defined the start and end date for the download, i.e., 01/01/2015 to 31/12/2019. The temporal resolution imported will vary depending on the temporal resolution of the data product. More information in each data product's section.

3. **Specify data product**. The image collection or data product was loaded. More information on the specific image collection used for each data product can be found in Table and described in detail in each data product specific section.

4. **Cloud-cover mask**. All the imagery I used comes from optical sensors. These cannot penetrate clouds. Thus, it is important to detect areas with thick layers of clouds and mask them from the dataset for their poor quality. This step is only required for satellite imagery, not for satellite-derived data products, as these have no cloud problem. Fortunately, MODIS has a cloud mask built-in. Landsat, on the other hand, does not. For Landsat imagery, GEE provides their own cloud scoring algorithm (*ee.Algorithms.Landsat.simpleCloudScore()*). This algorithm scores each pixel in an image based on its cloudiness. It then adds a band named 'cloud' to the input image which contains the exact cloud score ranging from 0 (absence) to 100 (cloudy). The user can then create a mask to flag those pixels which score is above a predefined threshold. A commonly used threshold is 20. Accordingly, I masked any pixels above a threshold of 20.

5. **Reduce spatial resolution**. For those data products that were provided at a resolution finer than our target resolution of $500m^2$; I performed a spatial reduction to match it. All images were re-projected to coordinate system: WGS84 (EPSG: 4326).

6. **Output export**. Images at the desired spatial and temporal resolution were exported as *.tif* files in designated folders in my Google Drive.

Identification and handling of missing data

Earth observation data gives us an opportunity to have highly resolved spatial and temporal information at a global scale. Nonetheless, missing values are abundant, most often due to miscellaneous phenomena that interfere in the measurement, such the cloud cover discussed above. This was the case for the features *bsa*, *sza*, *NDVI* and *LST*. There are several imputation techniques that can be used depending on the nature of the data and the distribution of the missing values. I first explored the existence of continuous gap-filled datasets already available and validated. This was the case for *lst*, for which a worldwide continuous gap-filled MODIS product exists, developed and maintained by Shiff et. al. 2021 [269]. For all other datasets, we used a *gapfill* imputation method proposed by Gerber et al., 2016[281]. Below we explore the approach we took for the different datasets.

(a) Worldwide continuous MODIS LST dataset, by Shiff et al., 2021 [269].
It consists on a data fusion method which combines MODIS data with a cloud-mask at 1km^2 with the Climate Forecast System Version 2 (CFSv2) modelled 2-m daily air temperature [282] . The MODIS dataset provides the detail and granularity on the temporal and spatial variability; the CFSv2 model provides the complete spatial coverage three times a day at the expense of a coarser spatial resolution (0.2°, approx. 20km^2 in the equator) and so, it can be used to borrow information to fill in the gaps for days/locations with missing data. The CFSv2 is NASA's fully coupled global NCEP Reanalysis model representing the interaction between the Earth's atmosphere, oceans, land, and sea ice from 1979 until present. It provides global coverage at a resolution of approximately 20km^2 with no missing values. The authors argue the use of the CFSv2 air temperature variable instead of the skin temperature on the basis data quality and accuracy, which is worst for skin temperature as numeric models fail to capture variations in the surface characteristics which affects more skin than air temperature.

The method is based on the premise that the temperature experienced at a given time and location can be described as the combination of a long-term average temperature for that specific time of the year and location (seasonality or climatology) and the deviation from that average or expected value linked to the weather (temperature anomaly). For each pixel, the seasonality or climatology was derived using Temporal Fourier Analyses (TFA) for both the cloud-free MODIS LST product and the CFSv2 product. The TFA mathematically simplifies complex and noisy data, decomposing it in series of functions or harmonics with specific amplitude and frequency. In this case, the TFA was defined to describe the seasonal cycles of temperature in terms of annual, bi-annual and tri-annual components (or 'harmonics'). These Fourier harmonics may be recombined, providing a smoothed signal, which is regarded here as the climatological expected temperatures ($LST_{clim}(t)$). The temperature anomaly ($LST_{anom}(t)$)for each day and pixel is then calculated as the difference between the observed temperature from the CFSv2 product and the MODIS LST climatological temperature. Finally, each missing pixel in the MODIS LST dataset ($LST_{cont}(t)$) is

replaced by the sum of the MODIS LST climatological temperature ($LST_{clim}(t)$) and the CFSv2 temperature anomaly ($LST_{anom}(t)$) (see Eq. 4.6). This values are then used to replace the cloud-affected pixels in the MODIS LST original dataset, creating a continuous gap-free dataset.

$$LST_{cont}(t) = LST_{clim}(t) + T_{anom}(t) \tag{4.6}$$

Interactive queries can be made on their web application powered by GEE, yet these have a temporal and spatial limit, far below the requirements of this project. Alternatively, the authors have included extensive documentation and all the code files, including those for validation, in their GitHub account. Using the GEE editor, the code can be used to generate continuous LST data for any time period and area of interest. Here, I used this code to generate gap-free continuous daily MODIS LST data for São Paulo study area. The global RMSE and Pearson's R coefficient for the entire study period and area were 2.34°C and 0.83, respectively (Table 4.2). For detailed documentation refer to Shiff et al., 2021 [269].

Table 4.2: **LST imputation performance metrics.** Annual and global RMSE and Pearson's R coefficient for the area under study.

Year	Pearson	RMSE (°C)
2015	0.83	2.37
2016	0.85	2.47
2017	0.83	2.31
2018	0.8	2.21
2019	0.86	2.34
Overall	0.83	2.34

A shortcoming of this method, pointed out and addressed recently by Zhang et al., (2022) [283], is the fact that missing values in the MODIS LST dataset are directly filled by those estimated from the CFSv2 model without contemplating the spatial continuity of the data. This can lead to boundary effects and a poorer model performance. They propose instead a spatio-temporal gap-filling framework to gap-fill missing values in the MODIS LST which seems to outperform Shiff et al.,'s approach with a global average root mean square error difference of -0.82°C between approaches. Unfortunately, the publication of said approach was subsequent to the development of my temperature model. Further research should evaluate the impact of using alternative LST dataset as an input feature in the model.

(b) Gapfill spatio-temporal interpolation method

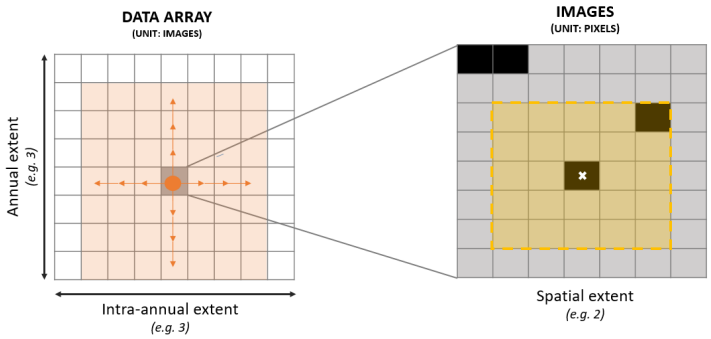
Available imputed datasets are rare. This was the case for the sza, bsa and NDVI datasets. So, I imputed the missing values using a novel spatio-temporal interpolation method called gapfill developed by Gerber et al., 2018 [281], which can be implemented in R software using the *Gapfill* package [284]. To ensure consistency, I applied the same imputation method across all three datasets.

Most data imputation methods exploit the temporal dependence of the data. However, remote

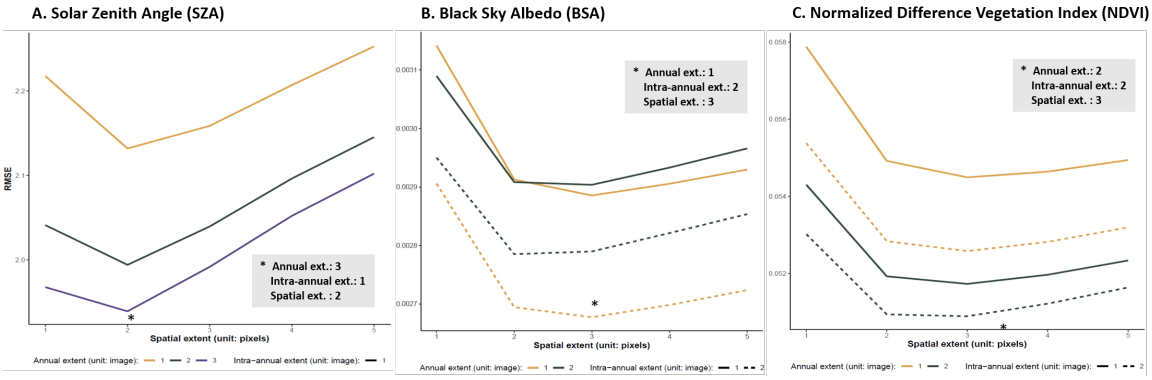
sensing data have both spatial and temporal dependence. The *gapfill* addresses this challenge and it does so at a low computational expense and providing uncertainty estimates. It has been shown to outperform TIMESAT and gapfill-MAP, both widely used temporal and spatio-temporal imputing methods. The *gapfill* method consists of four steps. Firstly, for any given missing pixel, one defines a prediction window (similar to a bounding box) containing the information to use for the imputation. Depending on the spatial and temporal dependence of the data, the size of the window will vary. Secondly, each pixel in the prediction window is given a similarity score based on all the other values in the window (performed pixel-wise). Images are then ranked in an increasing trend based on the score. Thirdly, for each pixel in the prediction window with a non-missing observation one determines the empirical quantile it corresponds relative to all values of the image. The mean of all the empirical quantile levels is then used as the target quantile level. Finally, all the quantile levels are then regressed using a quantile regression with level equal to the target quantile level. The target value can then be predicted based on the fitted quantile regression. An advantage of this method over others is that it provides statistical uncertainty estimates for each prediction. To do so, it estimates the uncertainty associated to each step, which is then combined to provide an overall uncertainty. For more details refer to Gerber et al., 2016[281].

The dimensions of the imputation window are key to the performance of the imputation. A window too small in size will cause the predictions to be noisy; a window too large and the predictions will be too smooth. This window is defined by an *annual, intra-annual* and *spatial extents* (Figure 4.13(a)). The *annual extent* defines the number of years to consider before and after to account for inter-annual variability. The *intra-annual extent* defines the number of images to consider before and after within one year. The units can vary depending on the temporal resolution of the data (e.g., daily, weekly). Finally, the *spatial extent* controls the number of pixels above/below and right/left considered during the imputation. To select the best parametrization for each dimension, I tested different window sizes using Cross-validation (CV) with a 30% random missing values. The accuracy was evaluated using RMSE. The window parametrization with the lowest RMSE was selected. Figure 4.13(b) shows the RMSE value of the parameters tested. The final window parametrization was (3,2,1) for *sza*, (1,2,3) for *bsa* and (2,2,3) for *NDVI*, for the annual, intra-annual and spatial extents, respectively.

Following the general process described above, I imputed the missing values in the *bsa*, *sza*, and *NDVI* datasets. Figure 4.14 shows the original dataset with missing values (top) and the imputed dataset for a randomly selected day of the *bsa*, *sza*, and *NDVI* datasets (bottom).



(a) Description



(b) Gapfill hyper-parameter tuning

Figure 4.13: **Gapfill prediction window hyper-parameter tuning** (a) Scheme of the components that define a *gapfill* imputation window.(b) Results of tuning of the prediction window components using cross-validation using a random sample of 30% missing values for (a) *sza*, (b) *bsa*, and (c) *NDVI*. Windows were defined by modifying the sizes of annual (color), intra-annual (dashed or solid lines) and spatial extent (*x-axis*). RMSE values as *y-axis*. Selected combination (lowest RMSE) indicated with an asterisk(*) and in the grey box.

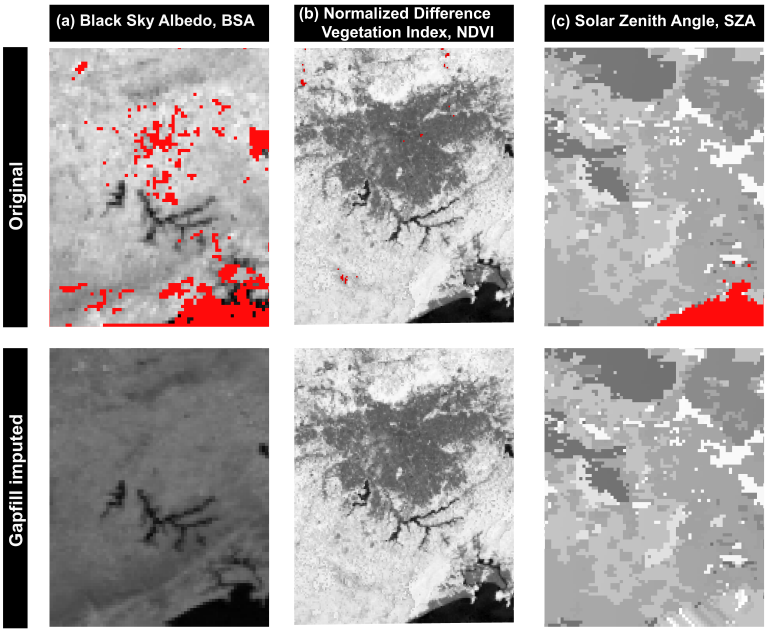


Figure 4.14: **Example of rasters imputed with *gapfill*.**Observed (top) versus imputed (bottom) for *bsa* (left), *NDVI* (middle) and *sza* (right).In red, the pixels imputed.

Temporal and spatial harmonization

Before building the model, I needed to re-sample all features to the target's spatial and temporal resolution ($500m^2$; daily). For the spatial re-sampling, I tested nearest neighbourhood interpolation and bilinear interpolation. The nearest neighbourhood interpolation algorithm assigns the value of the nearest cell without considering the surrounding values, yielding a piece-wise-constant interpolant. It is most used for categorical and ordinal variables, rendering a textured surface. The bilinear interpolation algorithm assigns a weighted average of surrounding cells, the four nearest ones by default. It produces much smoother maps and it is often used to re-sample continuous variables. I tested the use of both for each of the input features and selected the one that best captured its spatial variability without smoothing too much the extreme values, as these are important for the model training. For the temporal harmonization, I performed a cell-wise linear interpolation using as input the full raster stack available for the study period. The number of time slices available as input varied depending on the native temporal resolution of each dataset (e.g., every 3 months of NDVI, daily for LST).

4.2.4.3 Creating and preparing input dataset

I overlaid the monitoring stations with the raster stacks of the candidate input features and extracted the daily values that intersected each of the monitoring stations' locations. The resulting dataset contained a list all monitoring stations including their location and date of the recording (daily) and the values of each candidate input feature that intersected that location. To checked that the extracted values for each feature and station corresponded to the values for that day and input feature, I randomly selected two days and performed the join manually in QGIS. Since all candidate predictor variables were previously imputed, the only missing values that were considered here, were those from the monitoring stations (Figure 6.7). In total, I excluded 11% of the days (9,205 days out of 87,648).

4.2.4.4 Training and external validation dataset

To test the accuracy and performance of a model, I split the data in three datasets: training, testing and external validation datasets (Figure 4.15). First of all, I selected a sample of the original dataset to be the *external validation set*, also known as *hold-out*, which I put aside and never showed to the model training. This dataset must be representative of the original dataset and be able to represent the prediction scenario for which the model has been developed. In my case, the objective of the model is to predict at unsampled locations, hence, I selected a 10% random sample of stations to be the external validation. In order to ensure that they were representative, I iteratively selected 10% random stations and tested their representatives against all the data and selected the most representative split (Table 4.3). This data was then put aside until the end, when it was used to

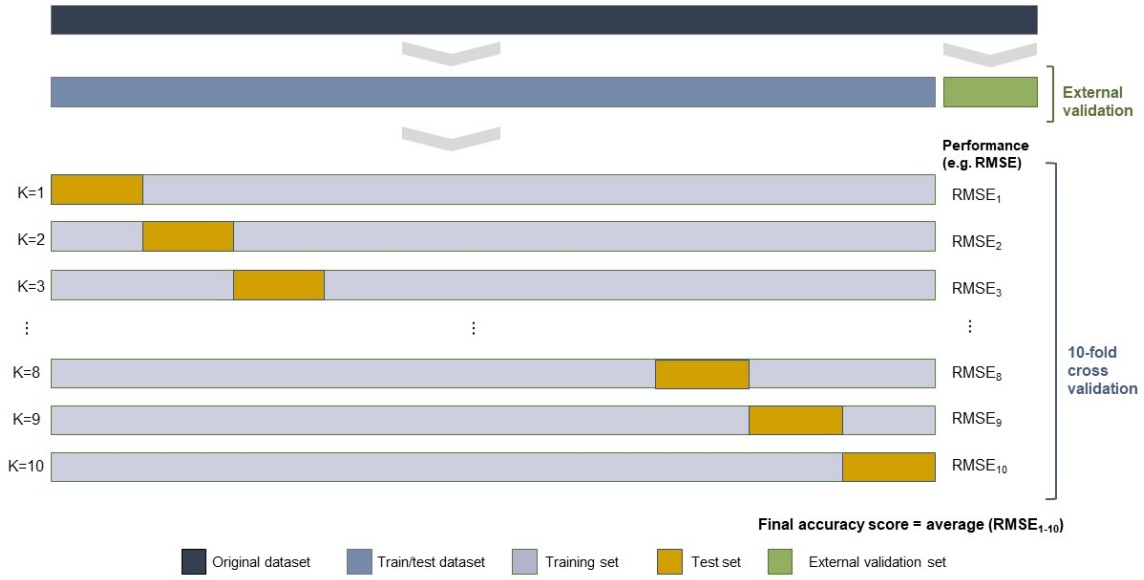


Figure 4.15: **Illustration of data split: train, test, and external validation.** Diagram of the data split between training, testing and external validation

validate the models once they were trained. For more detail on why having an external validation is desirable, please refer to the Section 4.2.6.2. With the remaining data, I iteratively split them in a test set and a training set. The training dataset was used to train the model and make the predictions; the test set was hold-out during the training to evaluate the performance the of the model in a process known as CV.For more details on how CV is performed and the different approaches that exist, please refer to Section 4.2.6.1. In short, in CV I iteratively train the model and test it on a hold-out sample of data to estimate the average performance. Depending on the approach used for the train/test split, I can distinguish different types of CV, each providing information on the accuracy and performance of our model under different scenarios.

Table 4.3: **Stations in training and external validation sets.** List of stations included in the training and external validation datasets based on random split

External validation	'A744', '1000944', 'CETESB_12', '1000840', 'CETESB_19'
Training set	'1000300' '1000842' '1000844' '1000848' '1000850' '1000852' '1000854', '1000857' '1000859' '1000860' '1000862' '1000864' '1000866' '1000876' '1000880' '1000882' '1000887' '400' '495' '503' '504' '507' '509' '510' '515' '524' '592' '634' '83075099999' '83779099999' '83780099999' 'A701', 'A755' 'A771' 'BR00E3-0520' 'CETESB_1' 'CETESB_18' 'CETESB_26' 'CETESB_6' 'CETESB_7' 'CETESB_9' 'SBGR' 'USP_station'

4.2.5 Model building

4.2.5.1 Random Forest (RF)

Random Forest (RF), the algorithm of choice for this modelling exercise, can be used for classification and regression tasks. Here, I provide an overview of the mathematical definition of RF for regression task. The goal is to build a predictive model that can predict the value of a target variable trained on a complete dataset containing the target variable and a set of features (training dataset).

RFs are supervised machine learning algorithms which operate as an ensemble of finite individual decision trees grown using a bootstrap sub-samples of the original dataset and applying feature randomness to create an uncorrelated forest of decision trees. Bagging refers to the use of bootstrap sub-samples of the original dataset to train the model. It was introduced by Breiman et al., 1996 [285]. For a dataset $\theta = (X_1, y_1), \ldots, (X_n, y_n)$ where n is the total number of observations, $X_{i(i=1)^n}$ the feature vectors or input auxiliary data, and y_n the outcome variable that I want to predict, θ_k represents sub-samples of the original dataset. Each sub-sample must be independent from the previous sub-samples $\theta_1, \ldots, \theta_{(k-1)}$ but have a similar distribution. While decision trees consider all the possible feature splits, random forests only select a subset of those features. This is known as feature randomness, and it generates a random selection of features with the aim to lower the correlation among decision trees and so, maximize the diversity between the groups. Finally, each sub-sample is trained independently (with feature randomness) and its predictions combined using an arithmetic mean to produce the final prediction of the RF. For a visual description refer to Chapter 2 Figure 2.5.

For an algorithm to learn, one needs to somehow estimate the information gain or loss at each iteration (i.e., learning rate). This is done through what in machine learning is called a *loss function*. This function evaluates how well a specific algorithm models a given data. When predictions deviate from the true values, the loss function increases, indicating a poor performance of the model. It is the responsibility of the optimization functions to minimize the loss. For a regression problem, the RF usually uses RMSE as the loss function (Eq. 4.7). At each decision tree node, the branch with the lowest RMSE, i.e., with the minimum loss, is favoured.

$$RMSE = \frac{\sum_{j=1}^{m}[\hat{T}(s_j, t_j) - T(s_j, t_j)]^2}{m} \tag{4.7}$$

where $\hat{T}(s_j, t_j)$ is the temperature predicted by the model at space s_j and time t_j; $T(s_j, t_j)$ is the temperature observed at that same time and space for a given RF node; m is the total number of observations for station j.

4.2.5.2 Hyper-parameter tuning and feature selection

A hyper-parameter is a model parameter whose value is defined by the researcher before the training process begins which influences the performance of the model. Often, for a given algorithm, some hyper-parameters will be set by expert opinion and others agnostically. Hyper-parameter tuning refers to the agnostic process of finding the set of hyper-parameters which minimizes the loss function. It involves defining a domain of hyper-parameter values to be searched, where each dimension represents a hyper-parameter with the scale equal to the range of values that hyper-parameter can take on, and so, each point a model configuration. There are as many dimensions as hyper-parameters need tuning. The ideal situation would be to test all possible combinations. However, this is exceptionally computationally intensive as the model needs to be fitted and its performance evaluated for each combination. Hence, researchers often use different strategies to test only a sub-sample of parameter combinations. Here, I used a grid search. According to this strategy, one defines a grid of n dimensions or hyper-parameters with x values to be tested and tests all the resulting combinations (nX). In this study, I tuned the following hyper-parameters: *n_estimators*, *max_features* and *max_depth* as they are some of the most sensitive to variability for model performance. Thus, my search domain had three dimensions (n=3). I then defined a set of possible values that each hyper-parameter could take which resulted in nine values (x=9) (Table 4.4). In total, I tested 27 combinations (nX = 3 X 9).

Table 4.4: **Hyper-parameter definitions.** Definition of the tuned parameters according to Scikit-learn *RandomForestRegressor* package. In bold, the values selected by hyper-parameter tuning.

Hyper-parameter	Description *	Values
max_features	The number of features to consider when looking for the best split	"auto":max_features=n_features; **"sqrt":max_features=sqrt(n_features)**
max_depth	The maximum depth of the tree	10; **15**; 20
n_estimator	The number of trees in the forest	300; 500; 700; **1000**

* Source: Scikit-learn RandomForestRegressor.

To further reduce the computational expense of this strategy, it is a common practice to use a sample of the original dataset for hyper-parameter tuning. Here, I used a 50% sample of the original dataset. To ensure that the sample was representative of the original dataset, I used *stratified random sampling*. Accordingly, I divided the original dataset into smaller groups using as a strata the target variable, *i.e., temp_mean75*, to ensure the sample had a similar distribution of temperature values. Figure 4.16 shows the distribution of daily mean temperature for the original (steel blue) and sampled data (ochre). To formally assess their similarity, I used a t-test (t-test=-0.1890; p-value = 0.8501) that confirmed the sample was representative of the population when it comes to the daily

mean temperature distribution, as suggested by the visual assessment. In bold in Table 4.4 are the hyper-parameter values selected based on the best model performance as per the grid search strategy.

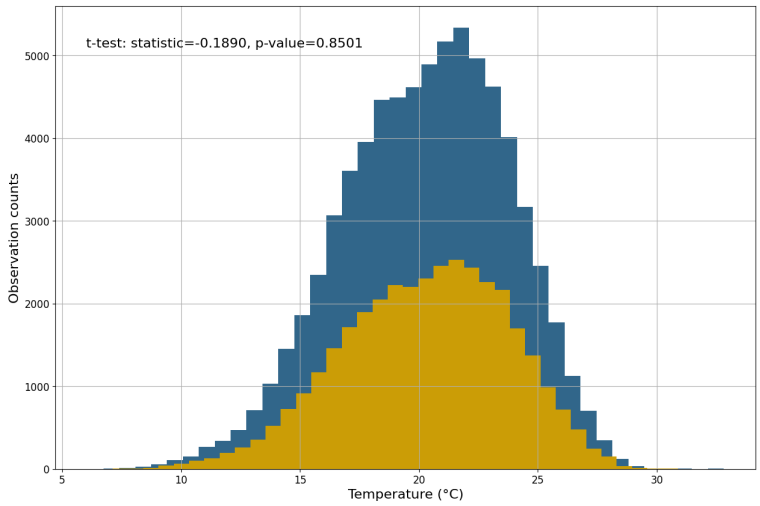


Figure 4.16: **Density plot of daily temperature for hyperparameter tuning** In blue, the original dataset and in ochre, the stratified random sample used for the hyperparameter tuning strategy. t-test statistic and p-value provided.

A common challenge in statistical modelling, and in particular with RFs is over-fitting. Over-fitting occurs when a model is too specific to the training data, which is likely to lead to poor generalization of the model to new predictor spaces. There are several reasons for over-fitting but a common one in spatio-temporal models is the inclusion of spatially or temporally static variables. These variables provide a poor representation of the predictor space –i.e., the sample of values is small and does not capture the full extent of variability present in the predictor space. This creates an issue of unique identity, either associated to the training locations in the case of temporally static predictors, or to time in the case of spatially static predictors. This causes the model to learn limited combinations of predictors which are specific to either the training locations (if temporally static variables are used) or to the time points (if spatially static variables are used). Therefore, to avoid over-fitting, it is important to carefully select the variables that provide useful information without causing too specific combinations that only apply to the training dataset. Meyer H, et. al. 2018 [286], argue that such variables can be automatically identified and removed using Forward Feature Selection (FFS) methods. The particularly of this feature selection strategy is that it uses target-oriented performance model evaluation instead of feature importance as traditional strategies do. The type of target-oriented CV used will depend on the final objective of the model. In my case, since the model was aimed at capturing the spatial variability of temperature, I used 10-fold spatial CV.

To perform FFS, I used the *sklearn SequentialFeatureSelector()* function which tests combinations of features to form a feature subset in a greedy fashion, that is, making the locally optimal choice at each stage. The algorithm starts with selecting the unique feature that provides the best model

performance. At each step, the algorithm tests the model performance after adding each of the remaining features and selects the best feature to add to the existing set. This process is repeated until all features have been tested. The best combination is that which minimizes the loss function, in our case RMSE. Following Meyer et al., 2018[286] suggestion, the loss function is estimated from a target-oriented cross-validation, in my case, a 10-fold CV. The FFS and hyper-parameter tuning are two key steps in the training process which are sensitive to changes on each others values. And so, I performed both process simultaneously. At the end, the selected (unranked) features were: *coast*, *sza*, *lst*, *rh*, *d2m*, *t2m*, *v10*,*u10* and *daylength*, and the selected hyper-parameters were: *n_estimators*=1000, *max_features*=sqrt, and *max_depth*= 15. I provide an extensive description of how I got to this selection in Section 4.3.2.1.

4.2.5.3 Feature importance

A common statistic used to explore RF's model outputs is *feature importance*. This indicator provides information on how much each feature contributes to the model predictions. The default feature importance statistic is the *gini importance or mean decrease impurity indicator*. This indicator importance relies solely on the training subset. As it is based on the training dataset only, it has a tendency to overestimate the importance of temporally static and spatially dynamic features, for they create unique IDs which contain large information for predicting within the training dataset, yet little to no information for predicting outside of it. An alternative to the gini importance score is the *permutation-based feature importance* or *mean decrease accuracy score*. This method assesses the feature importance in the context of predicting at unknown spaces/times using the test dataset. A baseline model is first fitted to the test dataset. The algorithm then randomly re-shuffles the values from one of the features in the test dataset, passes the dataset to the model to obtain predictions and calculates the performance. The feature importance is estimated as the average difference between the benchmark or baseline score and the modified scored after re-shuffling, thus giving an idea of how important that variable is to the model performance. In order to ensure the estimates are stable, it is common practice to repeat this process several times. This approach provides more realistic feature importance rankings when predicting to unknown spaces and it is less sensitive to favouring temporally static features that can cause over-fitting. Thus, here I used the permutation-based feature importance. The models were tested against the external hold-out test dataset. The final score is the average cross ten iterations.

4.2.6 Model validation

4.2.6.1 Spatial cross-validation

Both hyper-parameter tuning and feature selection base their decision on the improvement of the performance of the model, and this is often evaluated through CV. Cross-validation refers to the statistical method of evaluating and comparing the performance of learning algorithms by dividing data into two segments: the training and validation datasets. Typically, multiple cut-offs or folds are iteratively used, so that each data point has a chance of being validated. Their behaviour and performance is highly sensitive to the approach used to define this datasets. Thus, different CV approaches provide different perspectives on the suitability and performance of a model for a specific use. The end objective of this model is to make temperature predictions that can be used in epidemiological studies of temperature-mortality to address the common challenge of spatial missclassification in the exposure assessment. Thus, I have a special interest in assessing the performance and accuracy of my model when it comes to predicting at locations that it has not seen. For this reason, I used, *10-fold spatial CV* approach. As opposed to the random and non-spatial CV approaches, here the hold-out includes entire station groups, which mimics better the objective of the model to predict at unsampled locations. The 10-fold, refers to the number of station groups used, in this case 10. This approach has been previously shown to outperform other more classical cross-validation approaches such as the random split or the non-spatial fold CV for model training and feature selection [286]. The metrics for evaluating the model performance are described in Section 4.2.6.3.

4.2.6.2 External validation

External validation refers to the action of testing the model on a set of data that the model has never seen, as opposed to CV where the model is tested a subset of data that is then used in the subsequent folds. External validation provides an unbiased evaluation of the model performance as it uses data never not shown to the model during training. The external validation dataset (or hold-out dataset) must be representative of the model's ultimate purpose. In my study, the model trains on a network of data points with complete daily temperature information over a given temporal window, and it is expected to generate predictions at locations that has never seen for the same temporal window. Therefore, the validation dataset should be able to test primarily the spatial predictive performance of the model. Accordingly, I selected a 10% random sample of meteorological stations (5 out of 44 total stations). The metrics for evaluating the model performance are described in Section 4.2.6.3.

4.2.6.3 Metrics for evaluating model accuracy

Throughout this chapter, I will use the *coefficient of determination* (R^2, Eq.4.8), and the *root-mean-square error* (RMSE, Eq.4.9), as measures of accuracy of the predictions to evaluate and compare the models. See below for their mathematical expressions:

$$R^2 = 1 - \frac{\sum (\hat{T}(s_j, t_j) - T(s_j, t_j))^2}{\sum (T(s_j, t_j) - \hat{T}(s_j, t_j))^2} \tag{4.8}$$

$$RMSE = \sqrt{\frac{\sum_{j=1}^{m} [\hat{T}(s_j, t_j) - T(s_j, t_j)]^2}{m}} \tag{4.9}$$

where $\hat{T}(s_j, t_j)$ is the temperature predicted by the model at space s_j and time t_j; $T(s_j, t_j)$ is the temperature observed at the same time and space; m is the total number of observations at station j.

Finally, to test the results for bias, I followed Kloog et al., 2014 [287] approach, and estimated the temporal and spatial error associated to my model measured through a regression (R^2). The *temporal error* was calculated by regressing the difference between the observed temperature at time t and space s, and the annual mean temperature observed, against the difference between the predicted temperature at time t and space s, and the annual mean temperature predicted. The *spatial error* was calculated by regressing the station-specific annual mean estimates in observed temperatures against the station-specific annual means from the predicted temperature.

The algebraic notation is as follows:

$$y_i = \beta_0 + \beta_1 x_i + e_i \tag{4.10}$$

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - y)^2} \tag{4.11}$$

Where i represents an observation in the total number of observations considered, and y_i and x_i vary depending on the error being estimated. For *temporal error* they are:

$$y_i = \Delta(T) = T(s,t)_l - \frac{\sum^n T(s,t)_l}{n} \tag{4.12}$$

$$x_i = \Delta(\hat{T}) = \hat{T}(s,t) - \frac{\sum^n \hat{T}(s,t)}{n} \tag{4.13}$$

and for the *spatial errors* they are:

$$y_l = \frac{\sum^n T(s,t)_l}{n} \tag{4.14}$$

$$x_l = \frac{\sum^n \hat{T}(s,t)_l}{n} \tag{4.15}$$

where $T(s,t)$ and $\hat{T}(s,t)$ represent the temperature at time t and space s observed and predicted, respectively, and n is the total number of observations within a year l.

4.2.7 Model comparison

4.2.7.1 Land use regression, LUR

In exposure science, multi-linear regression models are a popular approach to explain and predict spatial and temporal variability of environmental variables. Land Use Regression (LUR) are a particular sub-class of multi-linear regression model which became particularly popular for air pollution modelling. Their name comes from the fact that they collected information about characteristics of the area where the monitors where located, originally land use data, which was associated to the emission and dispersion of the pollutant, hence the name 'land use regression' models. Refer to Section 2.3.2.4 for an overview of this technique in contrast of other methods, including RF. Among one of the advantages of regression approaches over RF is their capacity to extrapolate outside of the training range. This is of particular interest for applications where extreme values are important, as it is the case for temperature-mortality studies. Therefore, I used this comparison to assess whether my RF model is substantially underestimating extreme values.

The model assumed the following general form:

$$\begin{gathered} Temp_{t,s} = \beta_0 + \beta_1 LST_{t,s} + \beta_2 daylength_{t,s} + \beta_3 sza_{t,s} + \beta_6 t2m_{t,s} + \\ \beta_5 rh_{t,s} + \beta_6 d2m_{t,s} + \beta_7 u10_{t,s} + \beta_8 v10_{t,s} \\ \text{for:} \quad t = 1, \ldots, T = 6,940; \, i = 1, \ldots, I = 40 \end{gathered} \tag{4.16}$$

where $Temp_{t,s}$ is the predicted temperature at time t and location s, β_0 is the intercept, β_{1-8} are the linear coefficients describing the association between temperature and the selected covariates, namely land surface temperature ($LST_{t,s}$), day length ($daylength_{t,s}$), solar zenith angle($sza_{t,s}$), , dew point and temperature at 2m ($d2m_{t,s}$ and $t2m_{t,s}$), and wind components ($v10_{t,s}$ and $u10_{t,s}$), respectively. The features included in the model where those selected for the RF approach. Forward feature selection on a vanilla RF is not only a common approach to select features for multi-linear regression models, but in our case, it also ensures comparability with the RF model.

4.2.7.2 Prediction and performance comparison

I compared the models based on two aspects: on the one hand, the temporal and spatial distribution of the predictions, and on the other, the model performance. To compare the spatial distribution, I mapped the daily mean temperature for both models averaged overall, by year and by month. In addition, I estimated the *delta temperature* ($\Delta Temp$), calculated as the difference between the predictions from the RF model (reference) and the LUR model. I performed these calculations at the pixel-level and for the same temporal resolutions listed above. Finally, I plotted the correlation between the pixel-level temperature predicted by the RF and LUR models, and the histogram distribution of the $\Delta Temp$ for the same temporal resolutions. To evaluate whether there were differences in the temperature predicted by each model over time, I examined the annual and monthly predictions using box plots. To formally test the differences, I used a chi-square for difference. With regards to the model performance, for each model, I calculated the RMSE and R^2 values using 10-fold spatial-CV and external validation with the same hold-out dataset.

4.2.8 Sensitivity analyses

I performed several sensitivity analyses to assess the robustness of the model to changes in the input features. More specifically, I fitted the model including the feature *coast* (SA_coast), excluding *d2m* (SA_d2m) and including *coast* and excluding *d2m* (SA_coast_d2m)(Table 4.10). The CV and external validation of each sensitivity analyses was performed using all data. However, predictions were only produced for 2019 to reduce computational expense.

Table 4.5: **Sensitivity analyses.** Description of all sensitivity analyses fitted to evaluate the model robustness. Performance (RMSE) assessed using spatial-CV.

Name	Description
Main	['lst','sza', 'rh','d2m', 't2m', 'v10','u10','daylength']
SA_coast	Main model including feature *coast*
SA_d2m	Main model excluding feature *d2m*
SA_coast_d2m	Main model including feature *coast* and excluding *d2m*

4.2.9 Data access disclaimer

Computer code for replicating the results will be posted upon publication of the research in my GitHub repository [25] and on the project's website [26].

4.3 Results

4.3.1 Descriptive statistics of the response variable and input features

The locations of the meteorological stations considered for this study are shown in Figure 4.1(right) as points if included in the model (n=43), and as triangles if excluded due to poor data quality or large proportion of missing data (n=17), colored by network. From hereafter, I will refer to the meteorological stations included in the model. No major issue was expected as data quality and coherence checks were already done during the pre-processing of the data (Section 4.2.4.1).

During the study period, the temperature ranged from 6.06 to 32.82°C with an average of 20.29°C (Figure 4.17 (b) and Table 4.6). In order to visualize the spatial variability, I mapped the annual mean temperature per station (Figure 4.17 (a)). From the map, one can see that there is some spatial gradient present, with temperatures differences of up to 2°C. In general, the highest temperatures were recorded in the built-up areas located in the center and north of the municipality's administrative boundaries; whereas the surrounding peri-urban and rural areas recorded lower temperatures. Figure 4.17(c) depicts the monthly mean temperature (red solid line) averaged across all stations during the study period. The shaded area, captures the maximum and minimum average monthly temperature registered across all included stations, providing a visual representation of the variability in monthly average temperatures across stations. This figures indicate that the spatial variability in temperature is present throughout the year, yet with slight temporal variations in the magnitude, reflected by the varying width of the shaded areas. The variability across stations remained when looking at weekly and daily average temperatures, yet the variability was larger (Appendix Figure 6.8). From Figure 4.17(c), it is clear that temperature followed a seasonal pattern, with the hottest months expanding from December to March and the coolest months expanding from June to August.

The list of predictors explored for the model training, along with details on their spatial and temporal resolutions and their data sources, is summarized in Table 4.1. All features were downloaded and pre-processed to ensure the spatial and temporal resolution matched the target, as described in Section 4.2.4. Figure 4.18 depicts the spatial distribution of each input feature over the entire prediction area. For those features with temporal variation, I randomly selected a day as an example (i.e., March 8^{th}, 2015). A daily visualization will be made available upon publication of the work in the project's website. For each feature, I confirmed that there were no *impossible values* (e.g., negative values in variable that only accepts positive integers); that all the values were within the expected range for each feature, area and time, and that their spatial and temporal variability followed the expected patterns. Table 4.6 provides summary statistics for each feature in the training dataset, calculated for the entire study period (2015-2019). Again, all values were within the expected range.

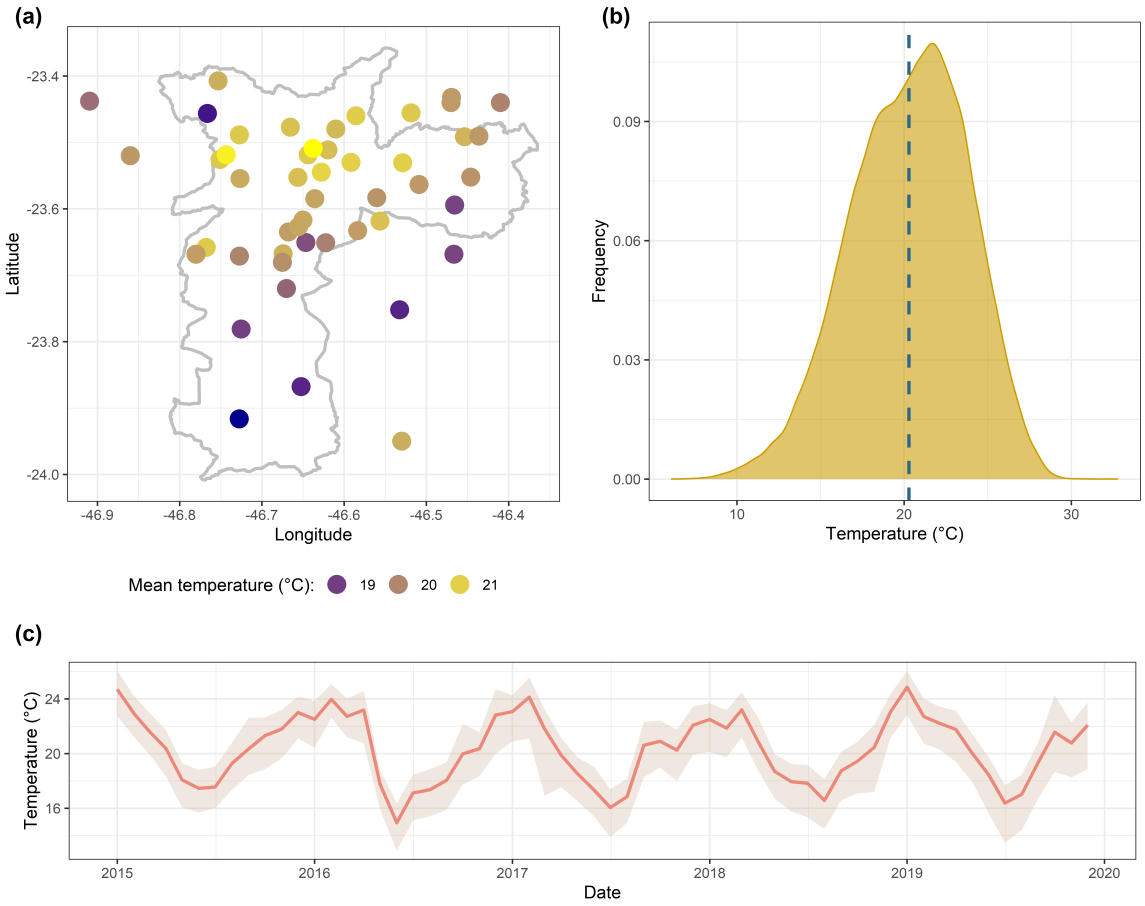


Figure 4.17: **Temperature from meteorological stations, spatial and temporal variability.** Daily mean temperature averaged over the study period (2015-2019) and shown (a) over a map of the study area by meteorological stations included in the model (n=48), (b) as an histogram with the overall mean shown as a dashed vertical blue line, and (c) over time as a solid red line. Shaded in pink the maximum and minimum monthly averages across station.

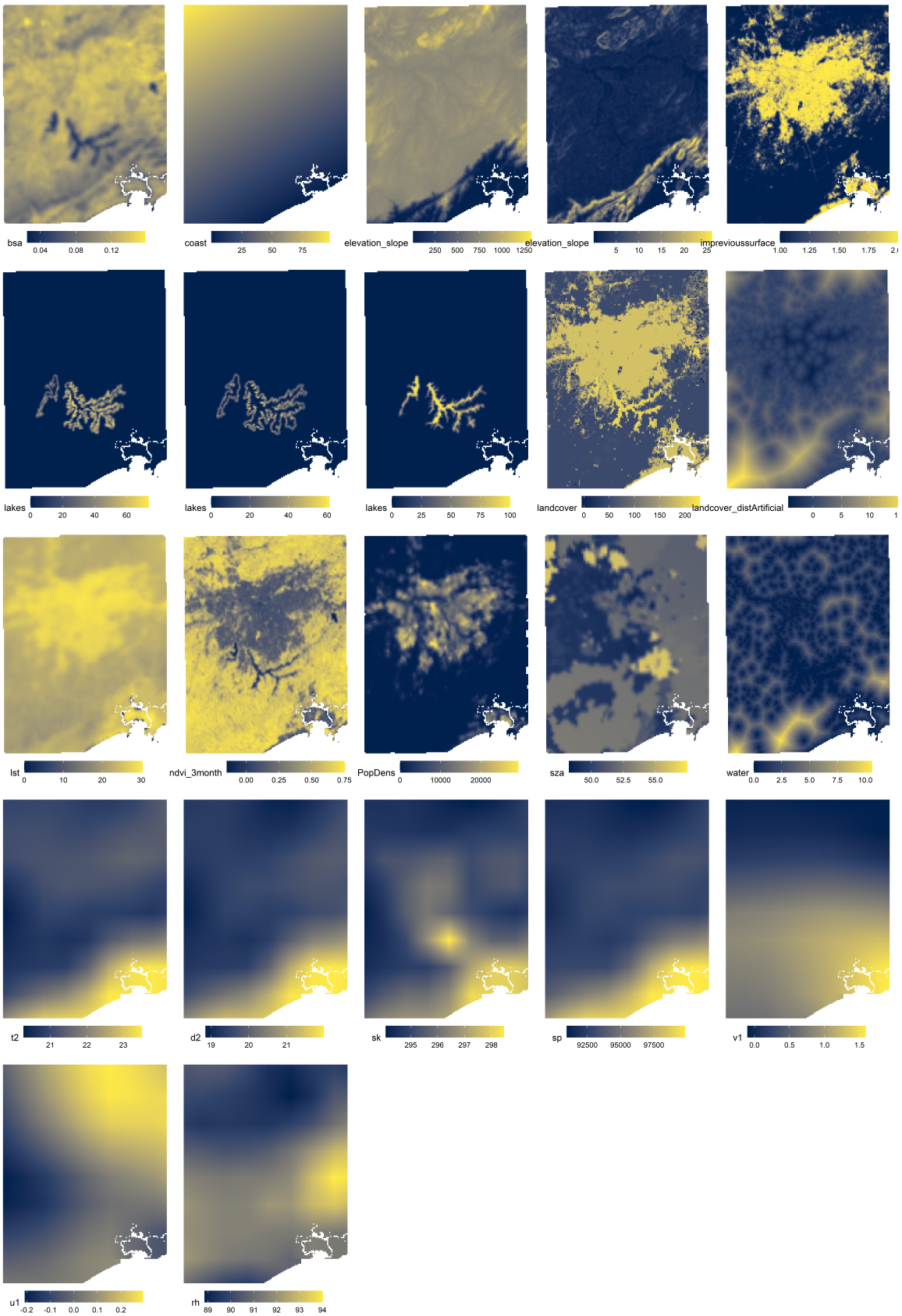


Figure 4.18: **Maps of the input features.** Map of the input feature at a 500m^2 spatial resolution for a randomly selected day (March 8th, 2015). Note: scales are specific to each map, please refer to the legend.

Table 4.6: **Descriptive statistics.** Descriptive statistics (minimum, 1^{st} quartile, median, mean, 3^{rd} quartile, and maximum) for observed temperature data and all the input features for the locations intersecting the meteorological stations included in the model.

Variables	Min.	1^{st} Qu.	Median	Mean	3^{rd} Qu.	Max.
Temperature	6.06	17.88	20.52	20.29	22.86	32.82
bsa	0.07	0.12	0.13	0.13	0.13	0.18
sza	13.60	26.87	34.85	36.20	45.74	56.39
t2m	7.34	17.57	20.03	19.80	22.19	29.28
d2m	3.18	13.87	16.72	16.28	19.03	25.36
NDVI	0.08	0.16	0.20	0.26	0.33	0.75
LST	8.40	22.03	25.08	24.86	27.92	38.47
coast	11.54	46.14	52.40	52.07	59.90	79.97
popdens	0.04	4183.70	8497.88	8398.40	12567.32	25703.31
impsurf	1.00	1.58	1.90	1.74	1.96	2.00
landcov	50.00	174.40	190.00	165.46	190.00	210.00
artland	-3.04	-1.42	-0.82	-0.57	0.04	4.90
lake	0.00	0.00	0.00	1.16	0.00	34.25
b200m	0.00	0.00	0.00	1.27	0.00	29.09
b400m	0.00	0.00	0.00	1.44	0.00	24.11
dem	370.85	747.49	761.91	760.47	782.41	991.41
slope	0.20	1.11	1.51	1.67	1.68	9.47
water	0.00	0.42	1.00	1.14	1.82	4.17
v10	-5.04	-0.79	0.33	0.23	1.32	4.55
u10	-4.56	-1.54	-0.67	-0.58	0.34	5.24
sp	91451.05	92755.57	93076.37	93165.11	93397.81	99088.97
skt	280.77	291.51	294.33	294.06	296.79	304.97
rh	33.42	76.16	82.39	80.65	86.78	98.66
daylength	10.65	11.12	12.08	12.10	13.08	13.62

Figure 4.19 shows the pairwise Pearson correlation coefficients amongst all input features in the training dataset. Below, I discuss some of the most remarkable associations and put them in context of what they mean for São Paulo. Impervious surfaces (*impsurf*) had a strong positive association (>0.50) with land cover (*landcov*, 0.87) and population density (*popdens*, 0.67). This is consistent with what I would expect as most population lives in densely urbanized areas which are primarily made up of impervious materials and that classify as 'build-up' in the land cover classification. This in turn, explains the strong negative correlation with the NDVI for all three variables (*impsurf:*-0.93, *landcover:*-0.62, *podens:*-0.87). The *daylength* variable captures the duration of sunlight, which is a proxy for seasonality as the duration of sunlight time varies during the year. Thus, *daylength* was positively correlated with black sky albedo (*bsa*, 0.63), land surface temperature (*lst*,0.68), and solar zenith angle (*sza*, 0.63). This behaviour was expected as *bsa* and *sza* are a direct consequence of the angle and intensity of the sun radiation which varies during the year. By definition, *lst* is the radiative skin temperature of the land derived from solar radiation and thus, longer duration of sunlight will lead to higher *lst* values. As expected, all inland water variables, i.e., *lake, b200m* and *b400m*, showed a high positive correlation as they capture different distance buffers to the same water body (refer to

Section 4.2.3.2).

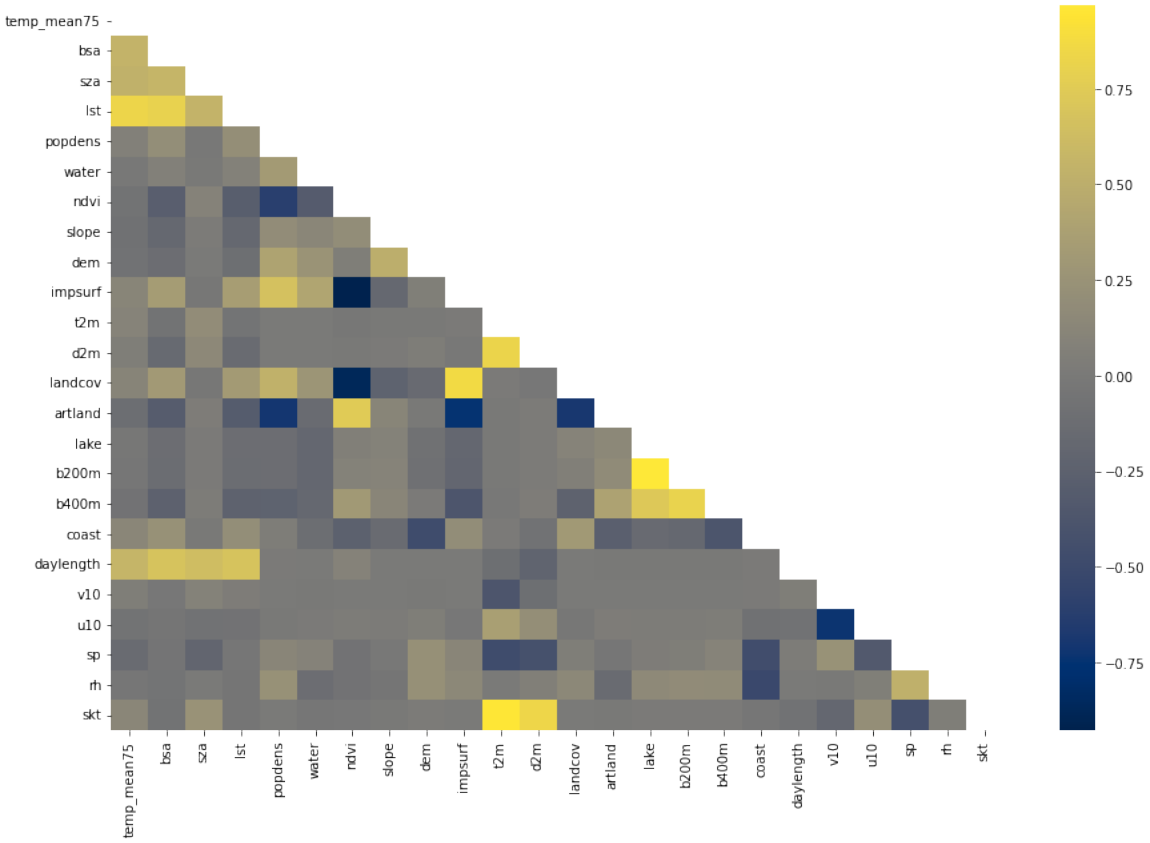


Figure 4.19: **Pearson correlation matrix input features.** Pearson correlation matrix for target variable and all input features.

The model relies on the existence of strong correlations between the input features and the target variable (i.e., daily mean temperature) to be able to make predictions. Figure 4.20 shows the Pearson correlation coefficients for each single-feature and daily mean temperature. This plots helped me understand if the associations between the input features and daily mean temperature follow the direction that I would expect. Of course, this is too simple of a score to capture complex associations such as non-linear associations and interactions, which may be present in the data. However, unexpected behaviours in the association at this stage can help flag potential issues with the feature input data. In our exploration, all the feature-temperature pairwise correlations behaved as expected. Thus, strong positive correlation were found for: land surface temperature (*lst*, 0.83), day length (0.56), black sky albedo (*bsa*, 0.55) and solar zenith angle (*sza*, 0.53). Negative correlations were found for day of the year (*doy*, -0.18), surface pressure (*sp*,-0.14), artificial land (*artland*,-0.12), slope and elevation (-0.9 and -0.8, respectively). The remaining variables behaved as expected yet the correlation coefficients were weaker (<0.2 for positive, and >-0.1 for negative correlation coefficients). All correlations were in line with what was expected based on the literature.

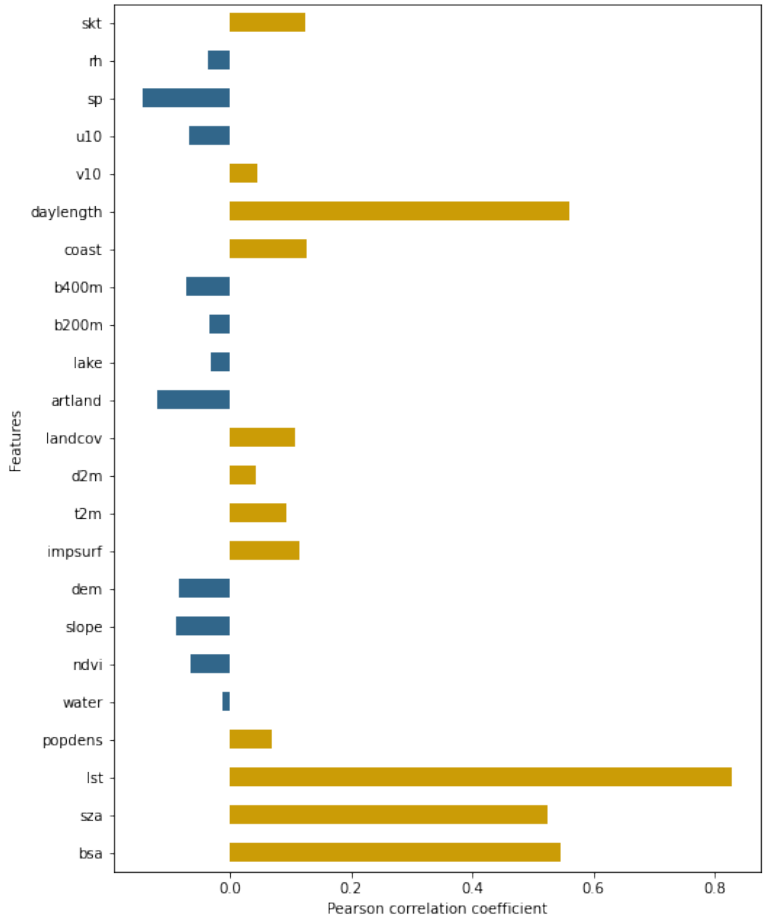


Figure 4.20: **Pairwise Pearson correlation with temperature** Bar plot showing the pairwise Pearson correlation coefficient between each input feature (*y-axis*) and the target variable (i.e. daily mean temperature). Positive correlation sin ochre and negative correlations in steel blue.

In order to examine the shape of the association, I plotted the daily mean temperature recorded by the stations against the input feature values for each location and estimated the best linear regressor (Figure 4.21). From this figure it is clear that the associations between temperature and some of the covariates is non-linear, which further supports the decision to use a RF to model it.

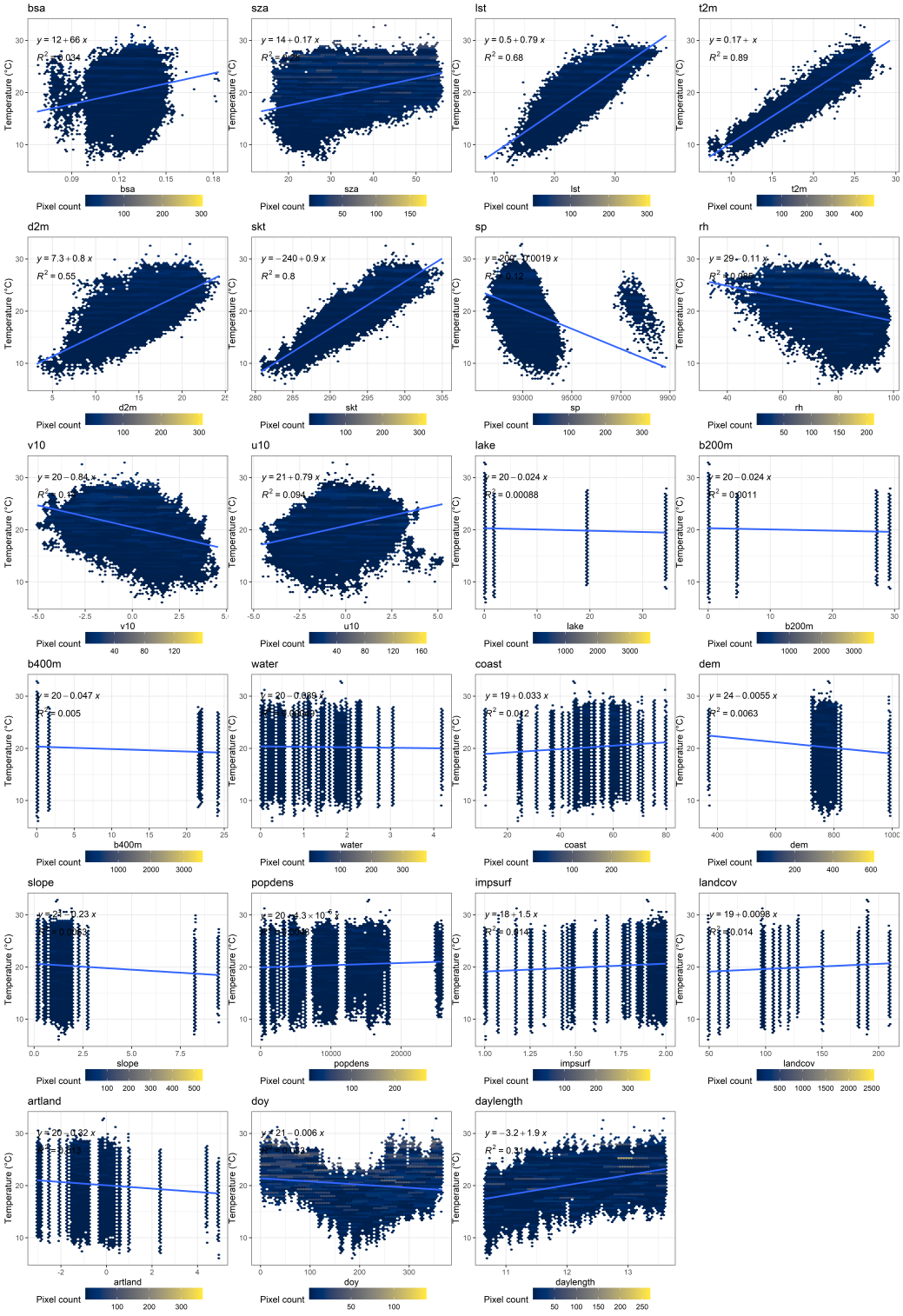


Figure 4.21: **Density scatter plot between daily temperature and input features.** Density plot of the daily mean temperature recorded by all included meteorological stations versus the values of the input features for those locations over time (2015-2019). The linear regression (blue line), R^2 and regression equation are shown for each plot.

4.3.2 Random Forest outputs

4.3.2.1 Model training

Figure 4.22 shows the data and modelling pipeline. Overall, the original dataset after doing the quality checks described in Section 4.2.4.1 contained 43 stations. The starting set of features was composed of 23 input features. After spatially and temporally joining them, the input data frame had a total of 87,648 observations. Of these, 10% of data was excluded due to missing values in the temperature recordings from the stations, resulting in a final complete dataset of 78,569 observations. Of the included stations, 10% (n=5) were hold out for external validation (n=6,799 observations), leaving a total of 43 stations and 71,770 observations for model training.

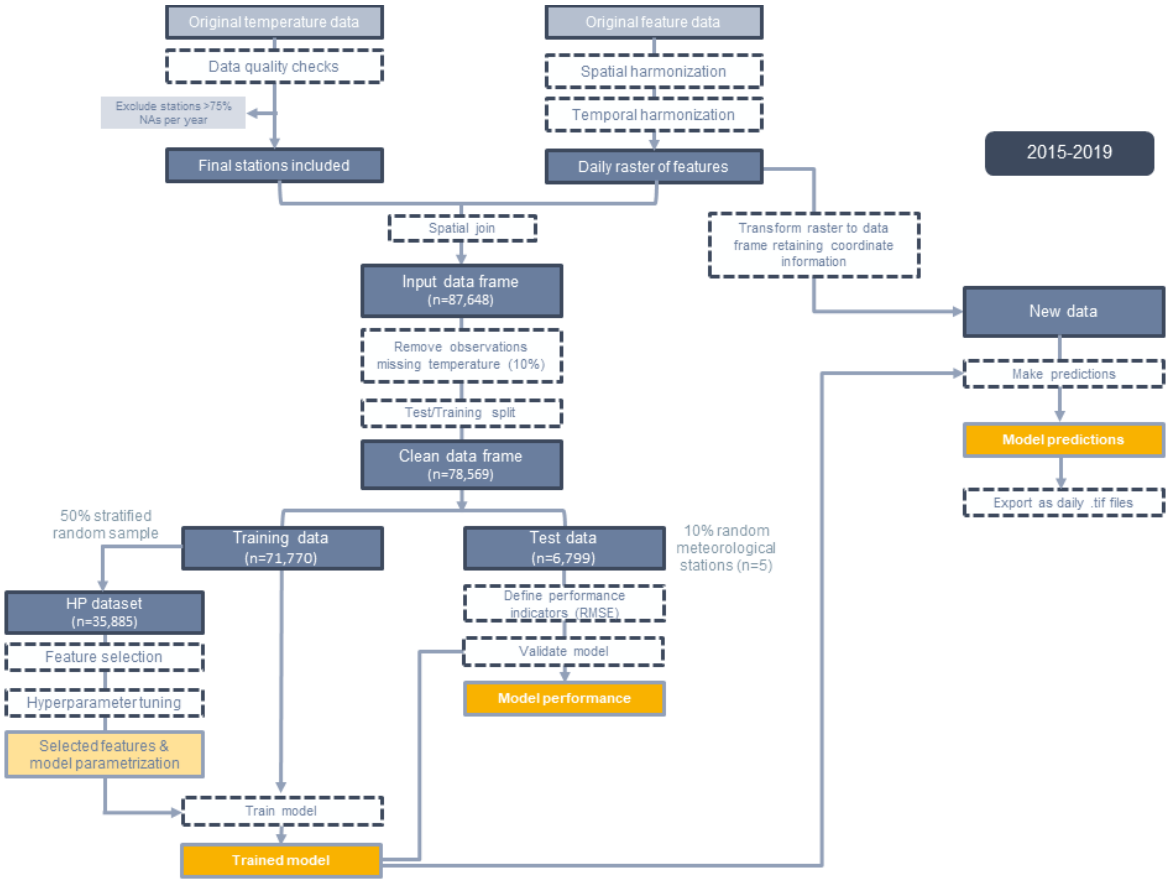


Figure 4.22: **Model workflow.** Diagram of the workflow followed, including data pre-processing, model training and predictions.

For the hyper-parameter tuning and forward feature selection, I used a 50% stratified random sample (n=35,885) representative of the overall dataset (t-test=-0.189, p-value=0.8501, Figure 4.16). Using a grid search approach, I run the FFS algorithm for each hyper-parameter combination (Table 4.4) and assessed the model performance for each. Figure 4.23 displays the distribution of RMSE resulting from each combination of features tested for each hyper-parameter combination. On the right, the results for a max_feature equal to *'auto'*, and on the left to *'sqrt'*. The number of trees in

the forest (*n_est*) is shown in different colours. RFs are computationally expensive to train with the expense raising with increasing complexity of the model. Thus, I defined the 'best hyper-parameter combination' as that which was the most parsimonious[3]. Looking at Figure 4.23, one can see that the best model performance is obtained at: *max_depth=15*, *n_est=1000* and *max_features= 'sqrt'*; thus, I selected these as the hyper-parameters for my model.

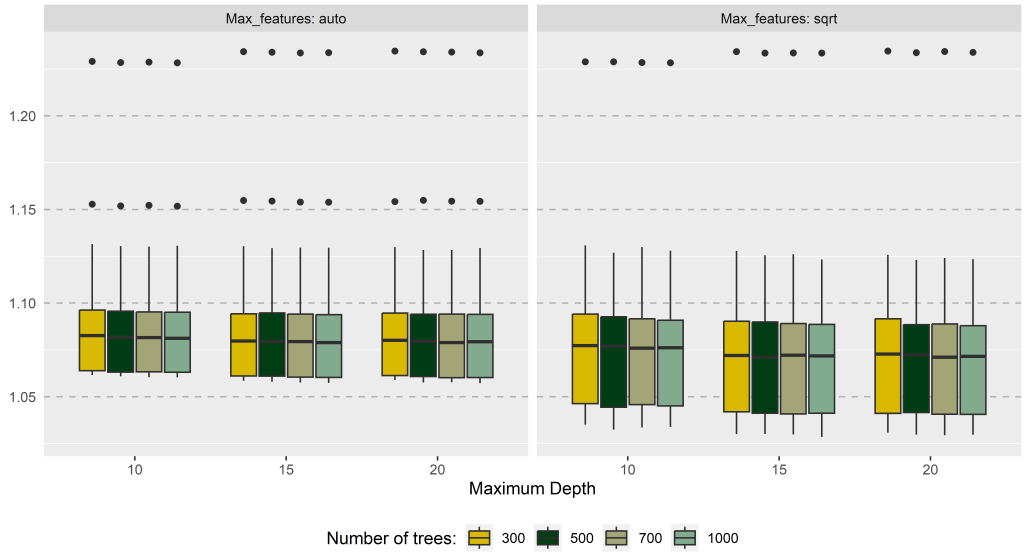


Figure 4.23: **Hyper-parameter tuning.** Model performance (RMSE) by maximum feature $(max_features, left and right hand side panels), number of trees($

Based on that hyper-parameter combination, I picked the feature combination with the lowest RMSE value, as that which had the overall lowest RMSE combination. Finally, I was interested in ensuring that the selection of features was robust. And so, I mapped the selected feature combination for the top 20 best performing models (regardless of hyperparameter combination). Figure 4.24 depicts for each hyper-parameter combination (rows) the features included (yellow) and excluded (blue) and the corresponding model performance (RMSE). At the bottom row, shaded, the best performing model. Another way to see this is presented in Appendix Figure 6.9, where each feature is assigned a score based on number of models of all tested hyper-parameter combinations (n=24) supporting their inclusion (yellow) or exclusion (blue). One can see that there is good accordance between the models, with most models selecting the same variables to be included/excluded of the model. There is only some heterogeneity in the decision for *coast* and *d2m* features with only half of the models (n=12) suggesting their inclusion. The inclusion of *t2m*, which is highly correlated to *d2m*, may explain why its inclusion is not fully supported by all models. Similarly, the inclusion of *rh*, would explain why the feature *coast* is excluded, as they are highly correlated. Since coast is a temporally static variable, which can cause issues with unique IDs, I excluded it. In conclusion, the selected

[3]The *parsimony principle* in statistics states that a simpler model with fewer parameters is favored over more complex models with more parameters, provided the models fit the data similarly well.

(unranked) features were: *sza*, *lst*, *rh*, *d2m*, *t2m*, *v10*,*u10* and *daylength*, and the selected hyper-parameters were: *n_estimators*=1000, *max_features*=sqrt, and *max_depth*= 15. Using this combination of hyper-parameters and features, I fitted the model to the complete dataset (71,770 observations) and validated the performance using 10-fold spatial CV and external validation with10% hold-out stations (n=5 stations). Finally, I used the algorithm to predict daily mean temperature at a 500m^2 resolution over São Paulo for every day between 2015 and 2019. Below, I provide a detailed description of the model outputs, performance and predictions.

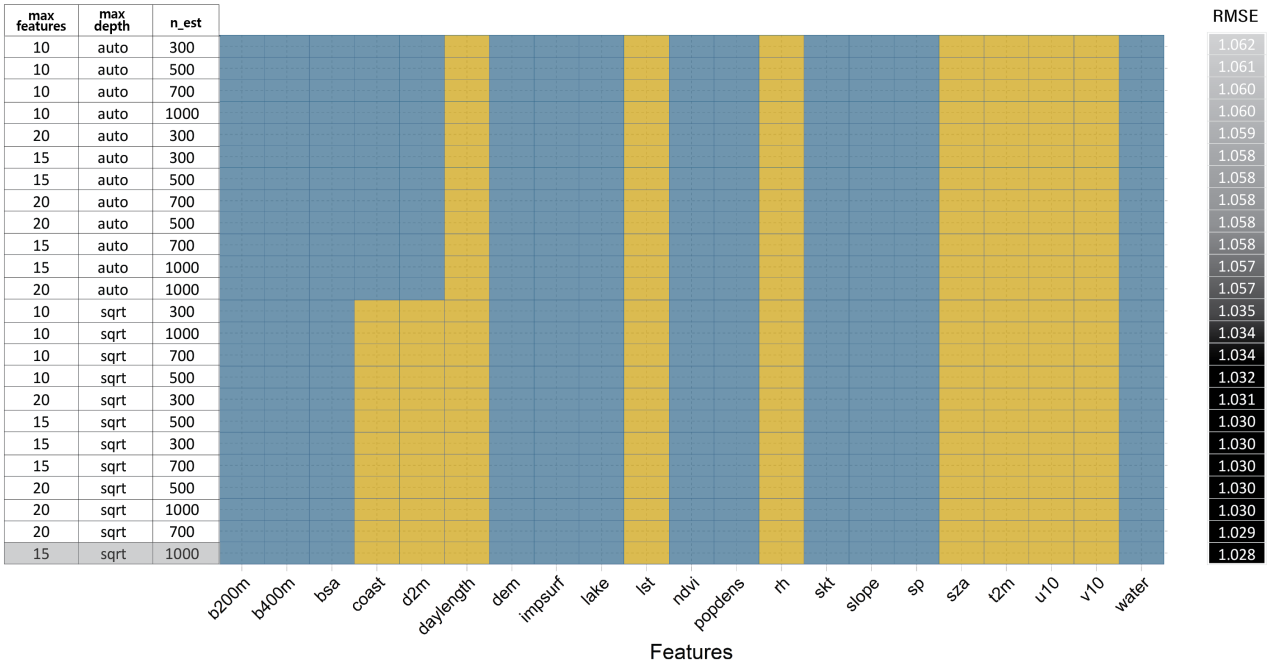


Figure 4.24: **Heatmap selected features per hyper-parameter combination.** (Left) Heatmap showing the feature combination resulting in the best model performance for each hyper-parameter combination. In yellow, the features selected and in blue those excluded. (Right) Performance values (RMSE) for each model are shown in a gray scale, sorted from top (highest) to bottom (lowest) by performance. Performance assessed using spatial cross-validation.

4.3.2.2 Feature importance

Figure 4.25 shows the mean importance score for the RF model according to the permutation-based feature importance approach. The temperature at 2m (*t2m*, 0.492) variable from ERA5-Land was the most important feature, followed by remote sensed LST (*lst*, 0.144), ERA5-Land dew-point temperature (*d2m*, 0.068) and RH (*rh*, 0.052), the ERA5-Land v-component of wind (*v10*, 0.028), and day length (*daylength*, 0.023). The variables which contributed the least to the model prediction accuracy were solar zenith angle (*sza*, 0.014) and ERA5-Land u-component of wind (*u10*, 0.011).

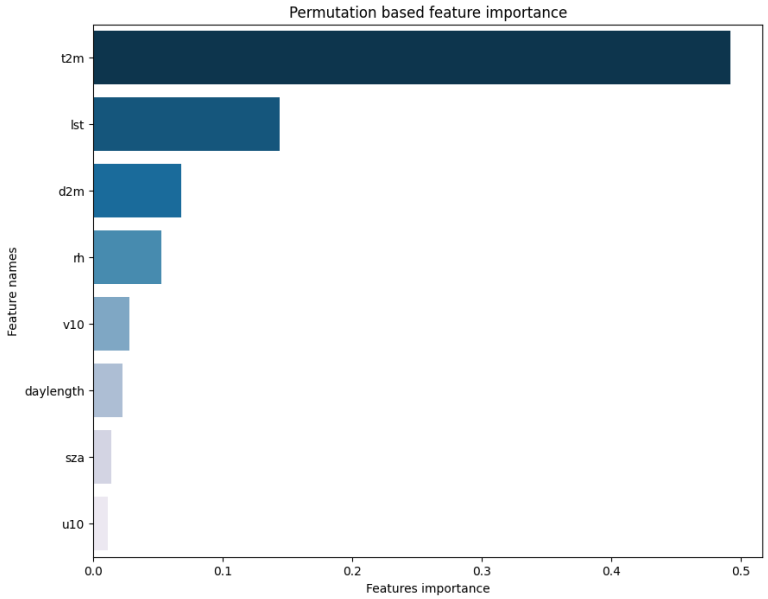


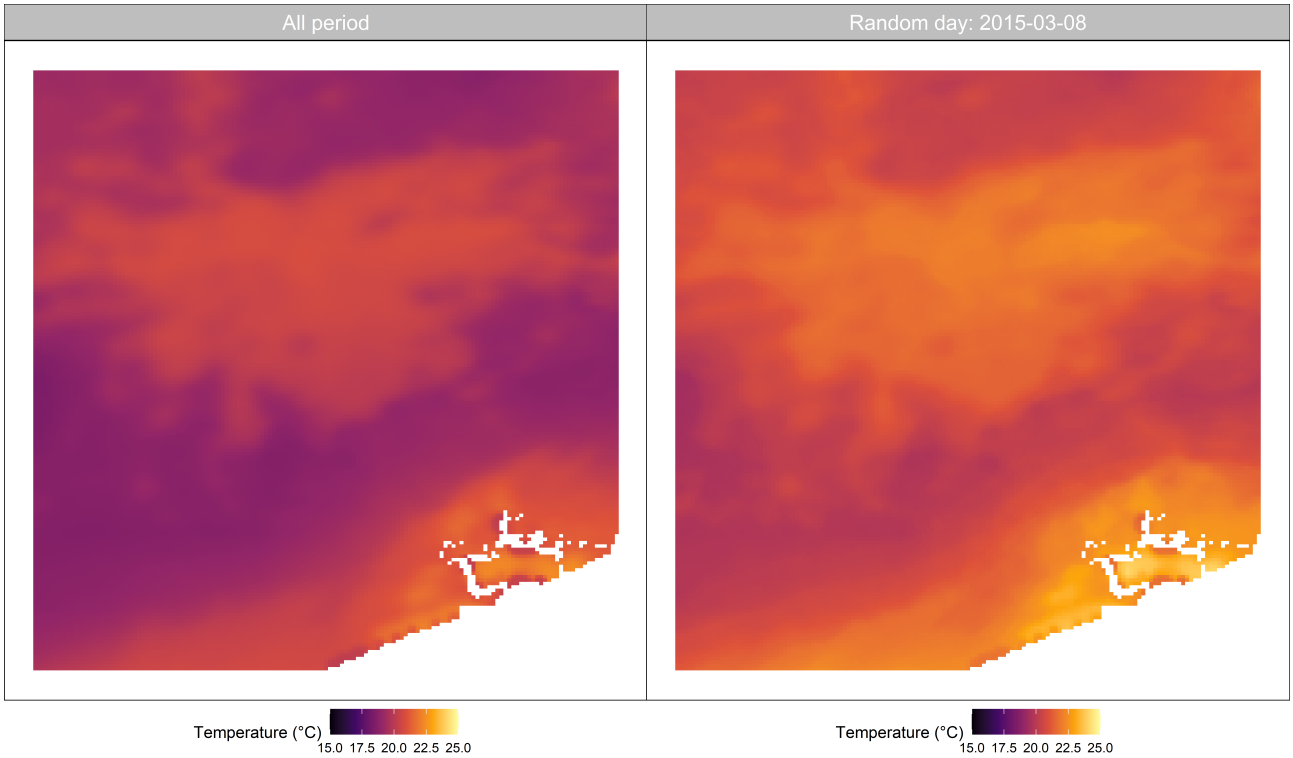
Figure 4.25: **RF feature scores per model** Permutation-based feature importance for Random Forest, RF. Scores representing the average across ten algorithm runs.

4.3.2.3 Model predictions

Spatial distribution of temperature

Figure 4.26 show the spatial distribution of the predicted temperature over São Paulo for all period (2015-2019) and for a randomly selected day (March 8^{th}, 2015) as an example. Apart from the obvious temporal variability that can be seen across the maps, which I describe in Section 4.3.2.3, there were remarkable spatial patterns. For example, there was evidence of the cooling and warming effects from land cover elements on the spatial distribution of temperature. In Figure 4.27, I have marked some of the land cover types present in the study area and provide on the left and right side an example of the landscape characteristic for each type. An example of the cooling influence of certain land covers is that seen for green spaces. The model seems to be able to capture the influence of both large and small vegetation coverage. Examples of large green spaces such as forests and natural reserves (dark green in Figure 4.27) are the *Parque Estadual Cantareira* located in the North of the city, and the *Serra do Mar*, which expands over the South East, following the coastline; examples of small greenspaces within the city such as parks (light green in Figure 4.27) include the *Parque Ecologico do Tiete*, *Carmo Park – Olavo Egydio Setúbal* and the *Parque de Ciência e Tecnologia da Universidade de São Paulo*. Regarding blue spaces, their effect on modulating temperature can also be seen, primarily, for the two dams located in the mid and Southern part of the study area (i.e., *Represa Billings* and *Represa de Guarapiranga*) (light blue in Figure 4.27). On the other hand, examples of warming effect of land cover types that the model seem to capture include mainly built-up areas. The effect of these areas, characterized by the large prevalence of impervious surfaces, high building and

population density and presence of anthropogenic sources of heat (yellow and beige in Figure 4.27), can be seen clearly with temperatures up to 5°C higher than the immediate rural surroundings. While the temperature differences were remarkable between rural-urban, difference between urban type are harder to distinguish and only perceptible at certain temporal resolutions. For example, there seems to be a tendency for temperatures to be slightly higher in the city center which is characterised by the presence of skyscrapers and abundant traffic (yellow in Figure 4.27). On the other hand, temperatures seem to be slightly lower in more residential and wealthy neighbourhoods (e.g., *Jardins*, *Itaim Bibi*, *Butanta*, *Perdizes*, *Pinheiros*, *Brooklin*) (beige in Figure 4.27). Slope and elevation, despite not being present in the model as input features, seem to be captured indirectly, most likely through the effect of *t2m* and *lst*. For example, I see that high mountain ranges such as those in the south of the study area, define a clear line in the predictions (grey in Figure 4.27). Interestingly, the rapid decrease in elevation (i.e., large slope) occurring across the *Serra do Mar* seems to have a remarkable impact, creating a sharp contrast on each side of the mountain range, concentrating warmer temperature on the seaside as the high mountains stop the warm breeze from penetrating the land.



(a) All period (b) Random day

Figure 4.26: **Map of temperature predictions for all period and random day.** Average daily mean temperature for (a) the entire study period (2015 to 2019) and for (b) March 8^{th}, 2015 (randomly selected day), as predicted by the Random Forest algorithm

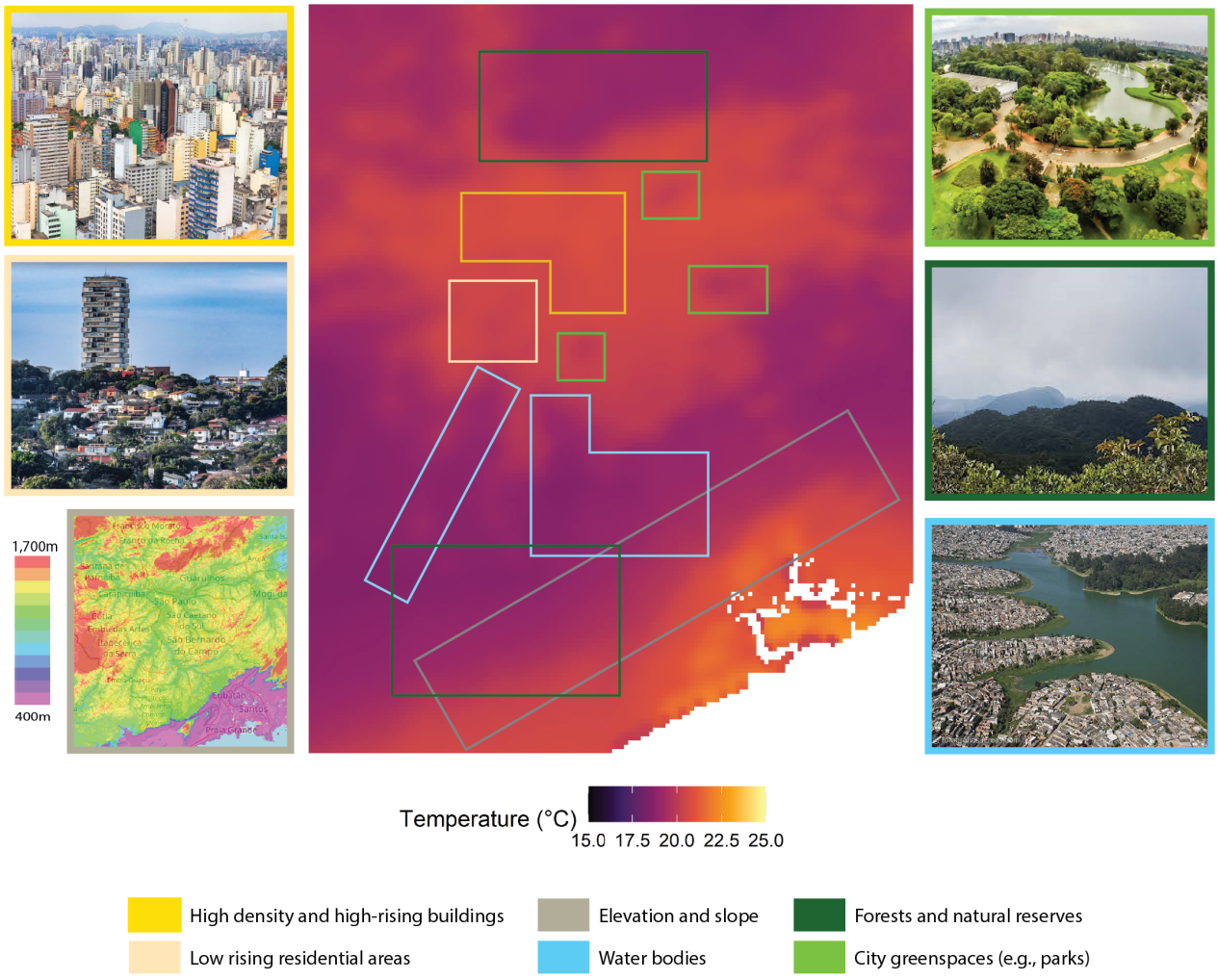


Figure 4.27: **Spatial distribution of predicted temperatures.** The image in the center shows the average daily mean temperature for the period 2015-2019. Colored boxes mark distinct spatial patterns in the temperature distribution. On the left and right side, exemplary images of the landscape characteristic for each box, with borders colored accordingly.

Temporal distribution of temperature

Figure 4.28 and 4.29 show the daily mean temperature averaged by year and month, respectively, as maps (a) and as a box plot (b). With regards to the annual variability, 2015 and 2019 seem to be generally slightly warmer years than the rest, with a mean annual temperature of 20.09°C and 20.02°C, respectively (Figure 4.28). Regarding the monthly variation, we see that the model is able to capture the seasonality of temperature, with the hottest months expanding from December to March and the coolest months expanding from May to August (Figure 4.29). This is in line with the seasonal oscillations observed for the ground truth (Figure 4.17).

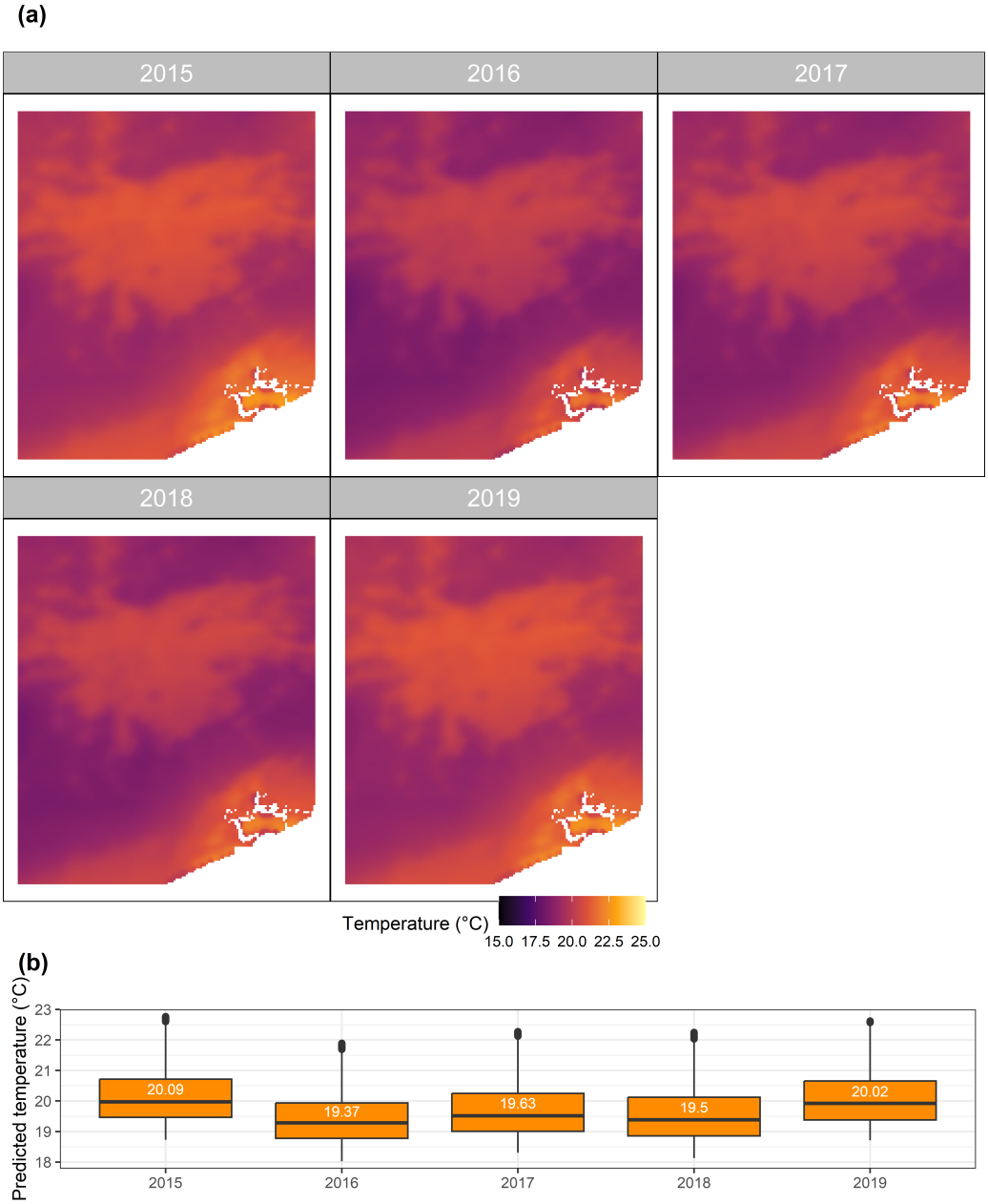


Figure 4.28: **RF temperature predictions by year.** Map of the daily mean temperature averaged by year across all study period (2015-2019) and (b) boxplot of the monthly temperature, as predicted by the Random Forest algorithm.

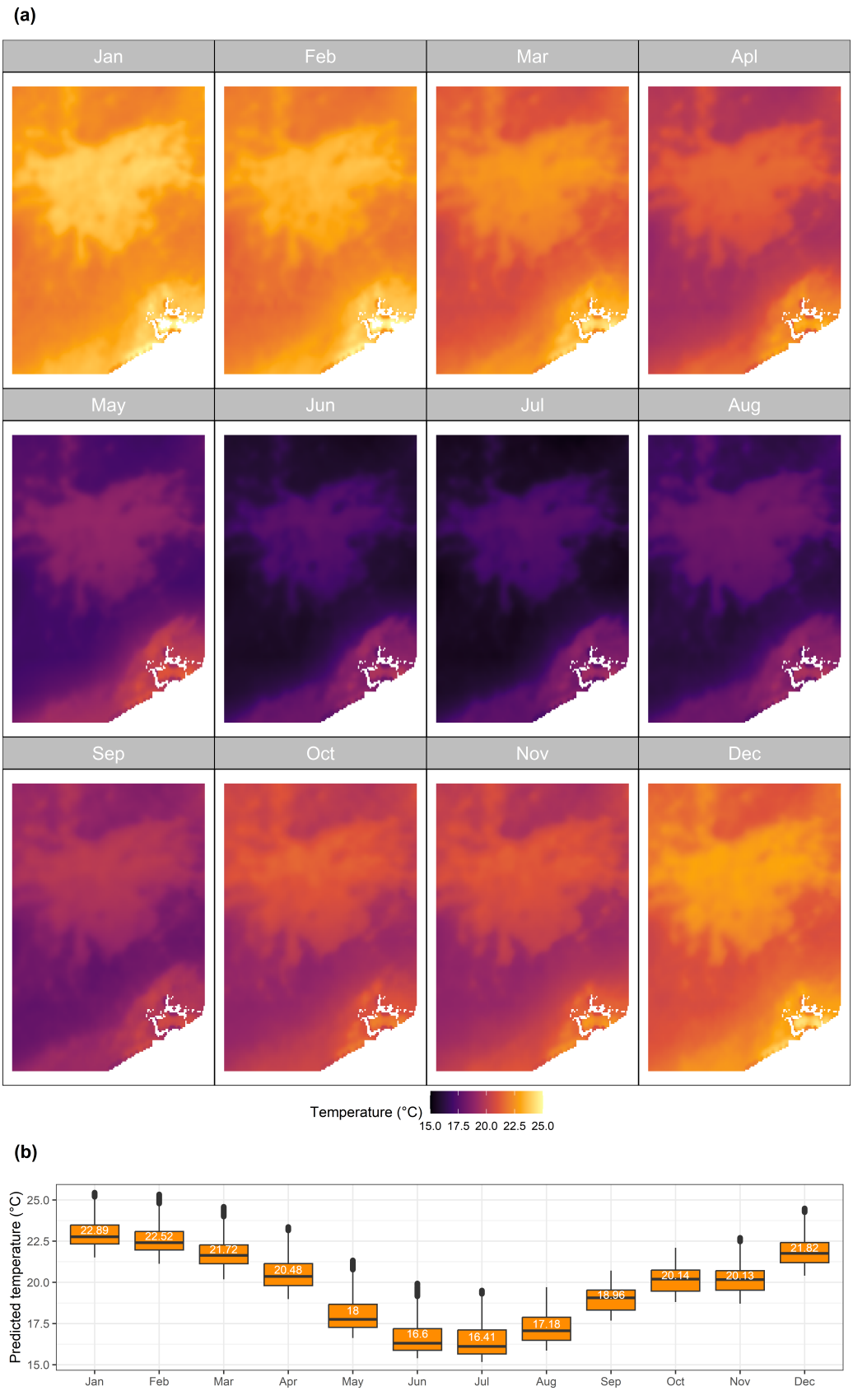


Figure 4.29: **RF temperature predictions by month.** (a) Map of the daily mean temperature averaged by month across all study period (2015-2019) and (b) boxplot of the monthly temperature, as predicted by the Random Forest algorithm.

4.3.2.4 Model validation

Spatial cross-validation

Using a 10-fold spatial CV approach, I predicted daily mean temperature at the locations of the included monitoring stations iteratively. Below, I explore the accuracy of the model through (i) scatter plots of the observed against the predicted values with their fitted linear regression and R^2 coefficients, and (ii) box plots of the difference between the temperature observed and predicted by the RF model ($\Delta Temp$) (Figures 4.30 to 4.35) . I present each of them for the entire study period, and at different temporal resolutions, i.e., daily, monthly and annually. Finally, I explore the spatial variability in the model accuracy between urban and rural locations. Table 4.7 provides the RMSE and R^2 overall, by station network, year, month and urbanicity.

Daily: The scatter plot in Figure 4.30 shows the relationship between monitored daily mean temperature and the RF predictions across all study period. In general, the model showed a good fit (R^2 =0.95). Similarly, the model error was low, with a RMSE=0.80°C for the entire study period (Table 4.7). From Figure 4.30, we can see that the RF slightly underestimates high temperatures and overestimates low ones. Figure 4.31 (a) shows the mean (solid blue line), minimum and maximum (shaded) daily difference between the observed and the predicted temperatures across all stations. From this figure, there does not seem to be a systematic error present over time, that is, the model tendency to over or underestimate temperature seems independent of time. As an example, I plotted the daily mean temperature recorded by the "1000862" meteorological station during 2015 (both station and year were selected randomly) and that estimated by the model at the same location (Figure 4.31 (b)). One can see that differences between observations and predictions are minor. From the example presented here, it seems like in 2015 at that location, the model generally captured the high temperatures but missed some of the temperature drops, predicting slightly higher temperatures.

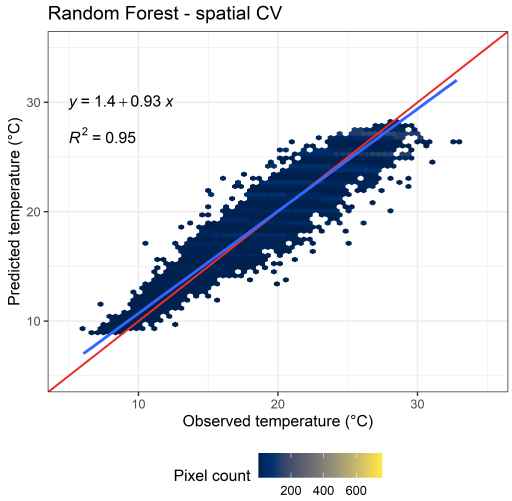
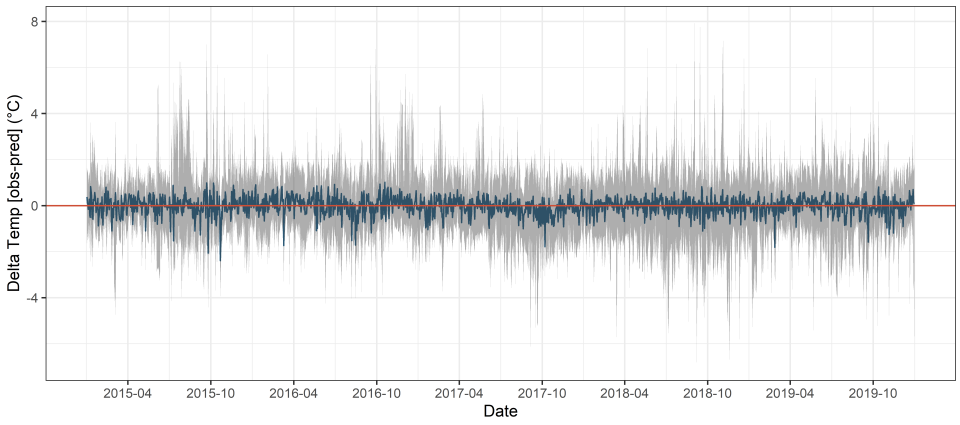
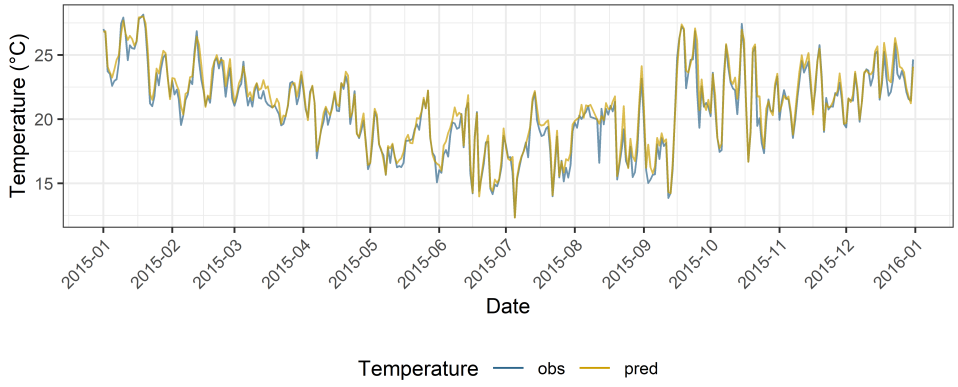


Figure 4.30: **Density scatter observed vs. RF predicted - all period.** Density scatter plots of daily mean temperature observed at the meteorological stations and that predicted by the RF model, for all study period (2015-2019), assessed through spatial-CV. The red and blue lines represent the 1:1 line and the linear regression, respectively. The R^2 and regression equation are shown for each plot.



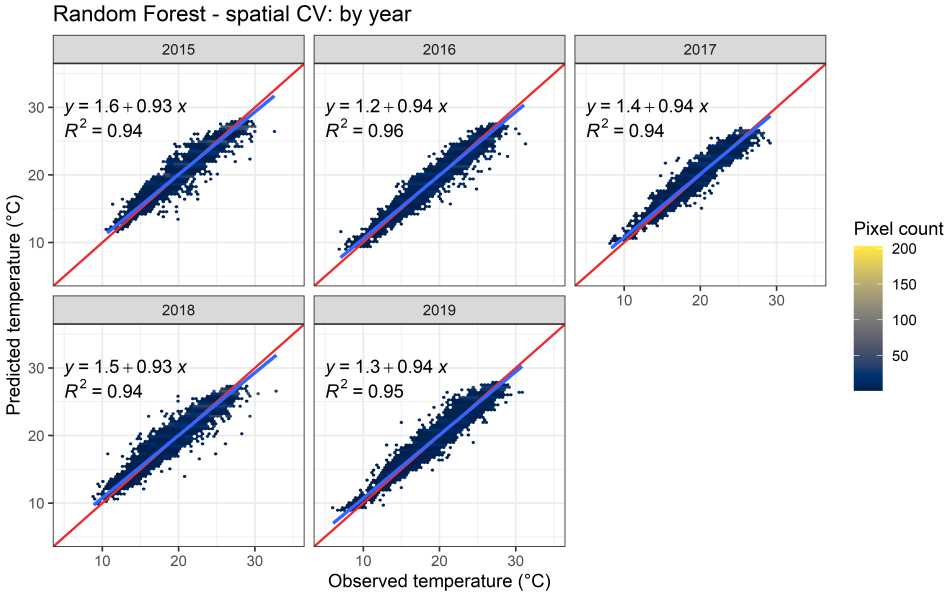
(a) Differences across all stations



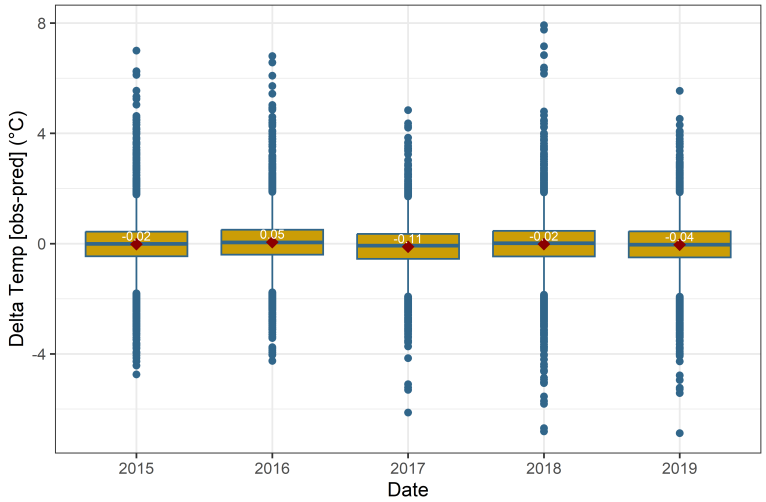
(b) Observed and predicted temperature at station 1000862 for 2015

Figure 4.31: **Difference between observed and predicted temperature.**(a) Mean (black), minimum and maximum (shaded) difference between observed and predicted daily temperatures across all stations for each day in the study period (2015-2019). The red line indicates the absence of differences (diff=0)(b) Predictions from the RF algorithm calculated through the spatial-CV (ochre) and observations (blue) for the 1000862 station for 2015 (randomly selected station and year).

Annually: The model accuracy remained rather stable across all years (Figure 4.32(a) and Table 4.7) with 2016 showing the highest accuracy (R^2 =0.96) and the second lowest error (RMSE=0.79°C). The box plots of temperature differences corroborate these findings, showing stable values across years (Figure 4.32(b)). The mean difference between observed and predicted temperatures across all years was -0.03°C with a SD=0.06°C, indicating little variation. However, it is important to note that in some rare occasions, the model predictions can be up to 8°C lower than expected (positive $\Delta Temp$) and up to 4°C higher than expected (negative $\Delta Temp$).



(a) Scatter plots for 10-fold spatial CV



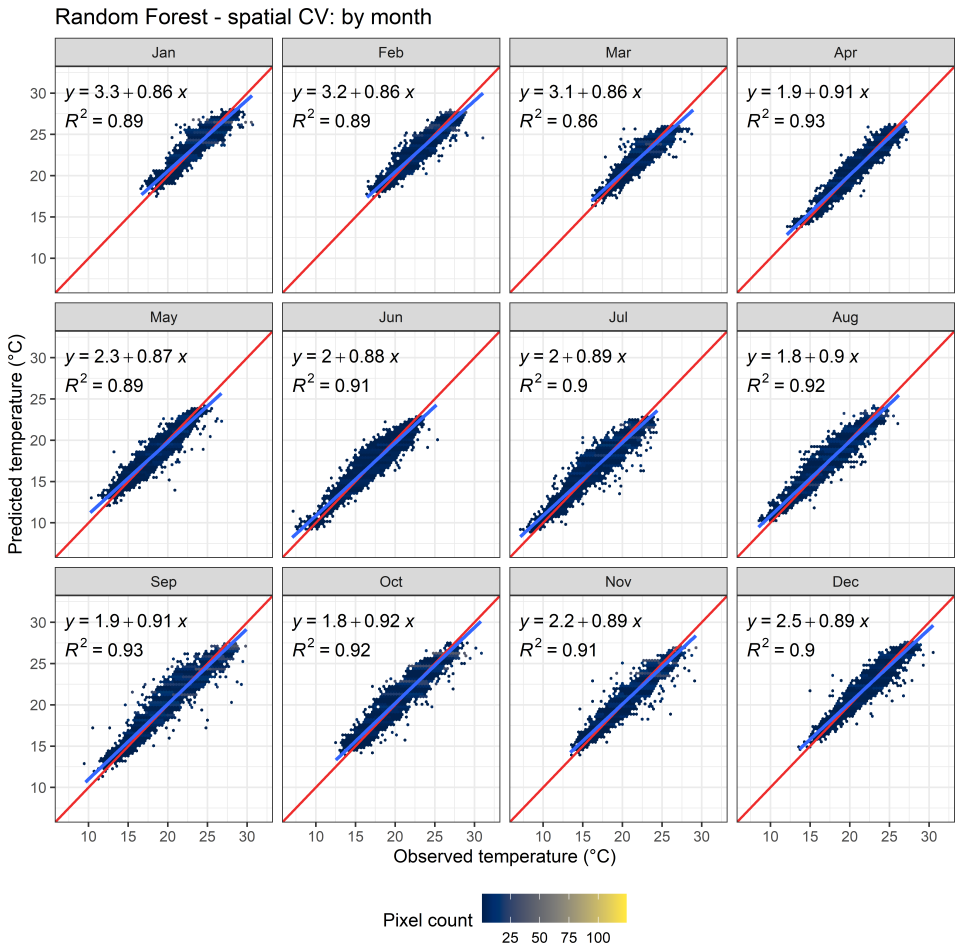
(b) $\Delta Temp$ box plot

Figure 4.32: **Annual differences in model accuracy and $\Delta Temp$.**(a)M Annual density scatter plots of predicted daily mean temperature versus daily mean temperature assessed through spatial-CV. The red and blue lines represent the 1:1 line and the linear regression, respectively. The R^2 and regression equation are shown for each plot. (b) Boxplot showing the difference between observed and predicted daily temperatures across all stations for each month in the study period (2015-2019). The red diamond indicates the mean; value label above.

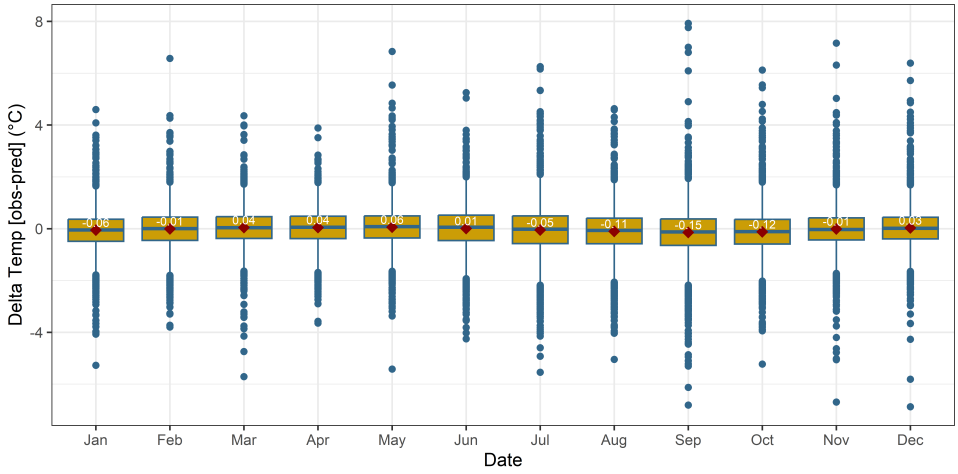
Monthly: When looking at monthly variations in model accuracy, one can see some more variation (Figure 4.33 and Table 4.7). In general, the accuracy of the model seems to be slightly worst for the hot months of January, February and March, with R^2<0.9. The higher RMSE values for those months corroborate this (Table 4.7) . The mean difference between observed and predicted values across all months was of -0.03°C, suggesting an almost negligible overestimation by the RF model. The variability across months was minor (SD=0.07°C) indicating that there was no systematic monthly bias in the model predictions. As described for the annual variation, it is important to state that in some rare occasions, the model predictions can be up to 8°C lower than expected (positive $\Delta Temp$) and up to 4°C higher than expected (negative $\Delta Temp$).

Spatially: Finally, I investigated whether there were any spatial patterns in the model accuracy. Figure 4.34 displays the RMSE per station. We can see that the lowest RMSE scores are concentrated in the center of the sample, which correlates with the location of all the urban meteorological stations. The same spatial distribution remained when stratified by year (Appendix Figure 6.11). Urban stations dominated the training dataset and so, I expected this to translate in an overall higher accuracy for those settings as it is corroborated by this findings. In order to further examine this rural/urban dichotomy in the model accuracy, I compared the model accuracy by urban-rural classification of the stations and found that the model performed better for urban settings (R^2 =0.95) as opposed to rural settings (R^2 =0.91)(Figure 4.35 and Table 4.7). To assess whether there was a systematic bias in the predictions between rural and urban stations, I plotted the observed difference by group (Figure 4.35b). From this figure, it seems like the model tends to slightly underestimate temperatures recorded by rural stations (mean=-0.22). In this case, despite they are still rare (outliers), we see from Figure 4.35(b) that the model underestimates temperatures with a difference of up to 8°C (positive $\Delta Temp$) and overestimate it with differences of up to 5°C (negative $\Delta Temp$) more often in rural than in urban settings.

Spatial and temporal error: Finally, I estimated the temporal and spatial error associated to my model measured through a regression R^2, as suggested by Kloog and colleagues 2014 [287]. Table 4.8 shows the annual R^2 and slope associated to the spatial and temporal component of the error. Overall, we can see that the overall temporal R^2 was 0.96 (year-to-year variation: 0.95-0.97) , much higher than the spatial one (R^2 =0.65; year-to-year variation: 0.58-0.69). This is in line with what I expected, as temperature has a strong seasonal pattern, which explains most of the variability and thus contributes to substantially to the overall good performance of the model. However, it is important to consider the poorer performance when regarding the spatial component. Finally, the slope values close to 1 indicate that there is little to no bias in my cross validation results, for neither the temporal nor the spatial components.



(a) Scatter plots for 10-fold spatial CV



(b) $\Delta Temp$ box plot

Figure 4.33: **Monthly differences in model accuracy and $\Delta Temp$.**(a)Monthly density scatter plots of predicted daily mean temperature versus daily mean temperature assessed through spatial CV. The red and blue lines represent the 1:1 line and the linear regression, respectively. The R^2 and regression equation are shown for each plot.(b) Boxplot showing the difference between observed and predicted daily temperatures across all stations for each year in the study period (2015-2019). The red diamond indicates the mean; value label above.

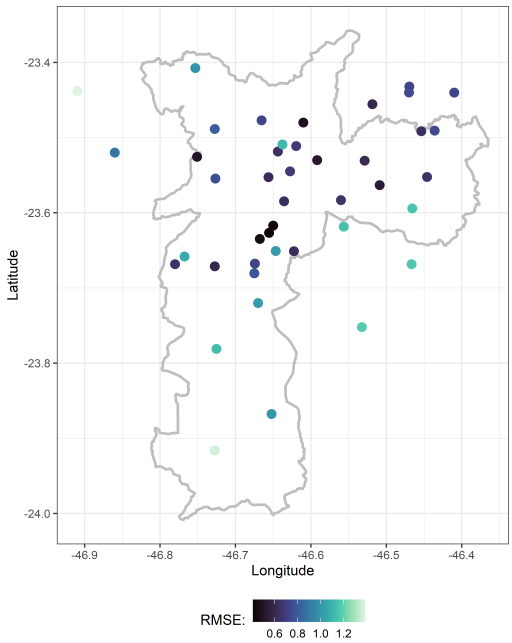
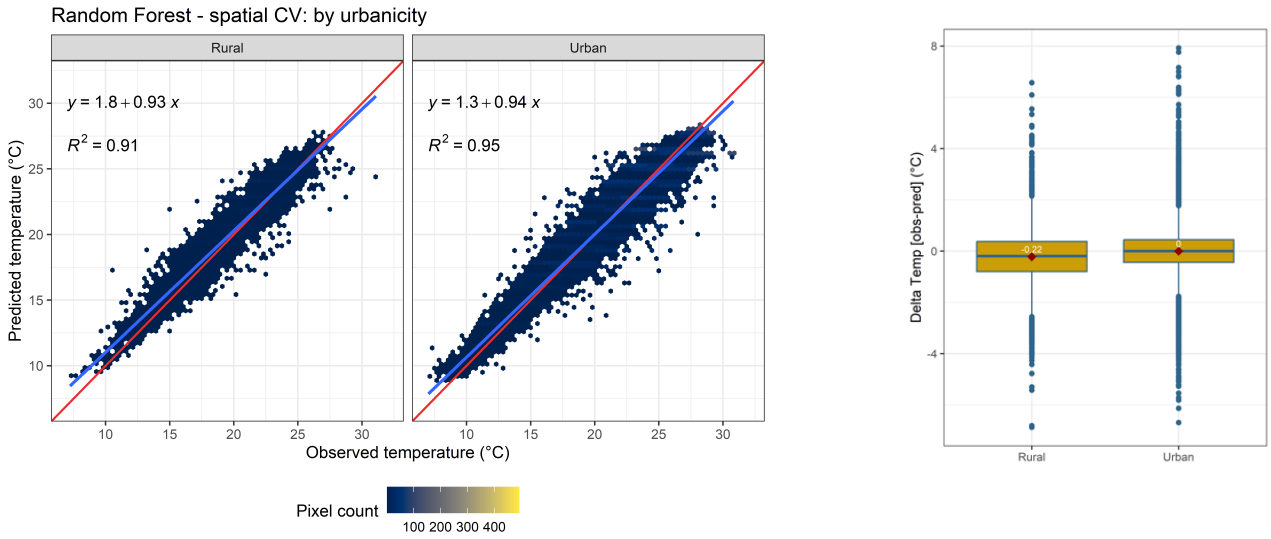


Figure 4.34: **RMSE per station for the period 2015-2019.** Average RMSE (°C) per station the period 2015-2019, assessed through spatial-CV.



(a) Scatter plots for 10-fold spatial CV

(b) $\Delta Temp$ box plot

Figure 4.35: **Rural/urban differences in model accuracy and $\Delta Temp$.**(a) Density scatter plots of predicted daily mean temperature versus daily mean temperature for stations classified as urban Vs. rural, assessed through spatial-CV. The red and blue lines represent the 1:1 line and the linear regression, respectively. The R^2 and regression equation are shown for each plot.(b) Boxplot showing the difference between observed and predicted daily temperatures ($\Delta Temp$) across all stations by urban/rural classification across all study period (2015-2019). The red diamond indicates the mean; value label above.

Table 4.7: **RF and LUR model accuracy.** Accuracy scores (RMSE in °C, and R^2) for the RF and LUR models for the whole study period and by meteorological station network, year, month and urbanicity, assessed through spatial-CV.

	Random Forest (RF)		Land Use Regression (LUR)	
Group	RMSE	R^2	RMSE	R^2
Total	0.80	0.95	1.02	0.92
airport	0.74	0.96	0.89	0.93
cetesb	0.73	0.96	0.90	0.93
cgs	0.84	0.94	1.06	0.91
ghcn noaa	0.43	0.98	0.77	0.95
gss noaa	0.80	0.95	1.05	0.91
inmet	0.76	0.95	0.86	0.93
USP	0.63	0.97	0.80	0.94
2015	0.80	0.94	1.04	0.90
2016	0.79	0.96	1.0v4	0.93
2017	0.78	0.94	1.03	0.91
2018	0.84	0.94	0.99	0.91
2019	0.81	0.95	0.99	0.92
Jan	0.74	0.89	0.94	0.82
Feb	0.74	0.89	0.92	0.83
Mar	0.69	0.86	0.84	0.80
Apr	0.69	0.93	0.87	0.89
May	0.75	0.89	0.92	0.84
Jun	0.84	0.91	1.08	0.85
Jul	0.94	0.90	1.09	0.86
Aug	0.86	0.92	1.11	0.87
Sep	0.94	0.93	1.30	0.89
Oct	0.85	0.92	1.09	0.88
Nov	0.76	0.91	0.95	0.87
Dec	0.75	0.90	0.95	0.86
Rural	1.05	0.91	1.15	0.89
Urban	0.76	0.95	1.00	0.92

Table 4.8: **Model accuracy by year.** Model accuracy by year (spatial and temporal component)

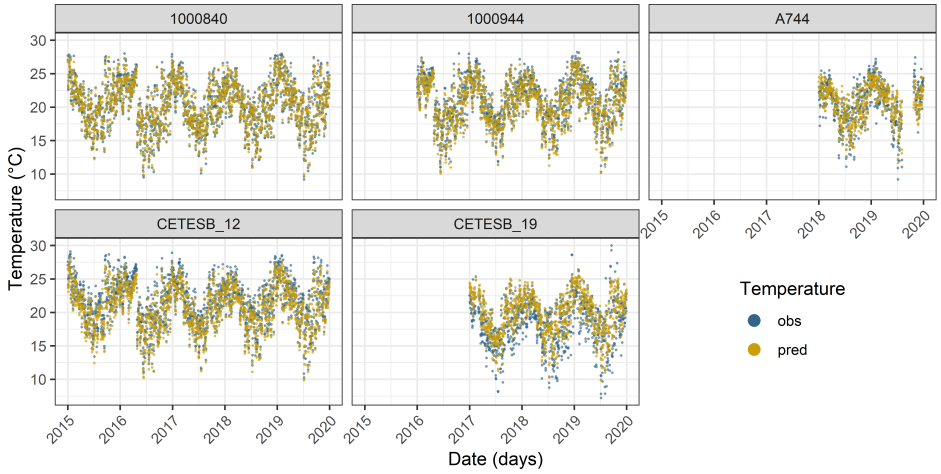
	Spatial component			Temporal component		
Year	R^2	Intercept	Slope	R^2	Intercept	Slope
2015	0.70	0.14	0.99	0.95	0.00	1.02
2016	0.64	0.54	0.98	0.97	0.00	1.02
2017	0.69	-1.40	1.06	0.96	0.00	1.01
2018	0.58	-0.06	1.00	0.96	0.00	1.01
2019	0.66	-0.25	1.01	0.96	0.00	1.01
Overall	0.65	-0.21	1.01	0.96	0.00	1.01

External validation

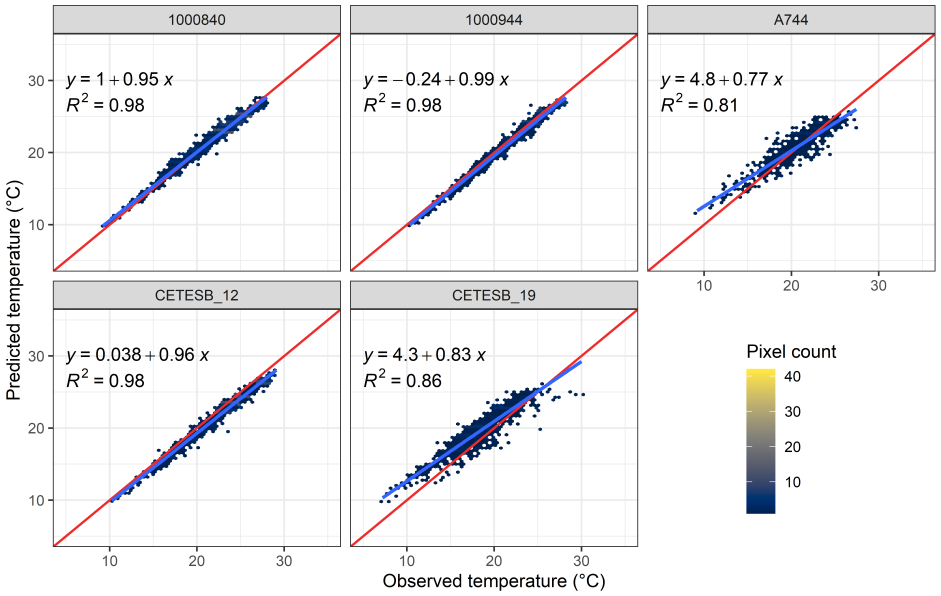
The overall accuracy across the five hold-out stations was rather high, with a RMSE of 1.00°C and an R^2 of 0.92 (Table 4.9). The accuracy of the model varied slightly by station, with R^2 ranging from 0.81 to 0.98. The worst performance was recorded for station CETESB_19 with a RMSE of 1.74°C and a R^2 of 0.86 (Table 4.9. Figure 4.36(a) shows the daily mean temperature predicted by the RF (ochre) and that observed (blue) in the five hold-out stations used for external validation. According to this figure, the predicted temperatures for the CETESB_19 station were systematically higher than the observed affecting primarily exterme cold temperatures. Together with the A744, CETESB_19 is the only other rural station of the hold-out dataset. Thus, these results are in line with the larger error and poorer accuracy observed in the spatial-CV results for rural stations.

Table 4.9: **External validation performance scores**. Performance scores (RMSE in °C, and R^2) calculated using the hold-out dataset consisting of 5 randomly selected stations. Results presented overall and for each station.

Group	RMSE	R^2
Total	1.00	0.92
1000840	0.47	0.98
1000944	0.66	0.98
A744	1.34	0.81
CETESB_12	0.90	0.98
CETESB_19	1.74	0.86



(a) Time series of RF external validation



(b) Scatter plots of RF external validation

Figure 4.36: **Time series and scatter plot of RF external validation.** (a) Time series plots of predicted daily mean temperature (ochre) versus daily mean temperature from the five meteorological stations (hold-out dataset, blue).(b) Density scatter plots of predicted daily mean temperature versus daily mean temperature from the five meteorological stations (hold-out dataset). The red and blue lines represent the 1:1 line and the linear regression, respectively.The R^2 and regression equation are shown for each plot.

4.3.3 Model comparison: LUR

In this section, I report on the results of the comparative analysis of the performance and prediction accuracy between the RF) model and the LUR model. The primary objective of this comparison encompassed two aspects. Firstly, it aimed to examine the robustness of the RF model in contrast to conventional regression techniques. Secondly, it sought to evaluate the extent of the challenge in predicting extreme values within our dataset, a well-recognized limitation of the RF methodology. To ensure comparability with the RF, I fitted the models with the same input feature data as that selected by the FFS during the RF model tuning (i.e., *sza*, *lst*, *rh*, *d2m*, *t2m*, *v10*,*u10* and *daylength*).

To evaluate the model assumptions and investigate whether or not there are outliers with a large influence on the analysis, I generated a histogram of the regression standardised residual, normal Q-Q plot and scatter plot of residuals against fitted values (Figure 4.37). From the figure, we see that the residuals tend to a normal distribution behaviour, with the classical bell shape in the histogram. The Q-Q plot suggest some deviance from normal at the extremes, which is expected. Finally, the lack of a systematic pattern in the distribution of the residuals across the fitted values supports the linear assumption, equal error variances and no problematic outliers. Thus, I proceed to discuss the outputs resulting from this model.

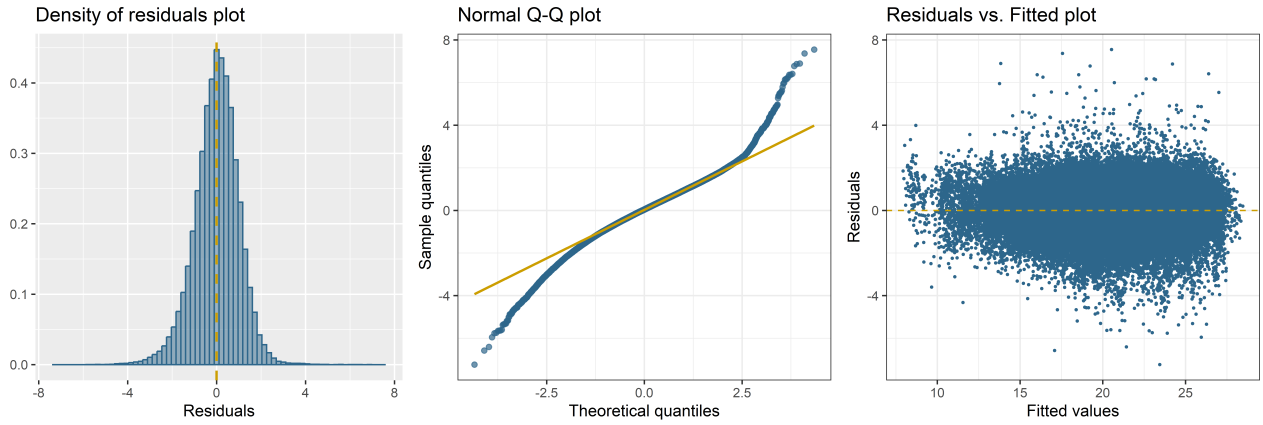


Figure 4.37: **LUR model residual diagnostic plots.** (Left) Density plot of the model residuals distribution. (Center) The Q-Q plot showing the empirically observed quantiles of model residuals (y-axis) as a function of the quantiles expected from a normal distribution with the same mean and variance as the empirical distribution (x-axis).(Right) Scatter plot of the model residuals plotted against the fitted values.

Feature importance comparison

Figure 4.38 shows the standardized regression coefficients of the LUR model. The associations followed the expected direction, with positive coefficients for *d2m*, *lst*, *t2m* and *sza* and negative coefficients for *rh*, *v10* and *u10*. The only coefficient with a unexpected value was that for *daylength* which showed a negative value. This can happen if the association is not linear and/or due to the presence

of interactions and effect modifications. From the sensitivity analyses, it seemed that *daylength* was relevant for the model performance, consistent with the variable selection in the RF model, and did not impact the predictions. Thus, it was kept in the model.

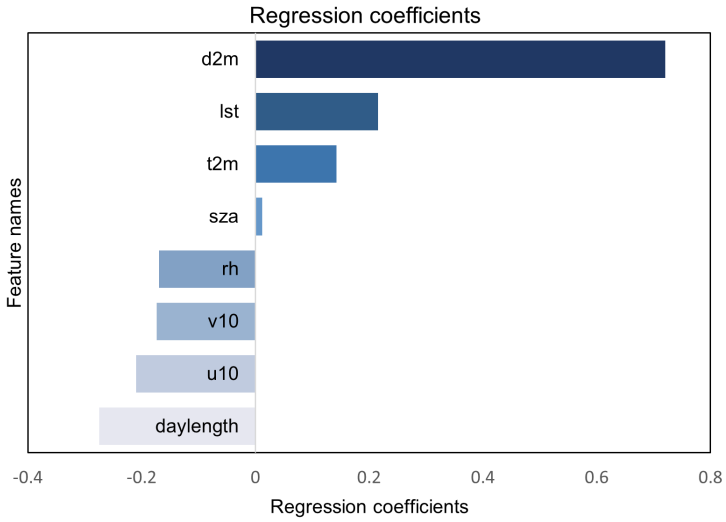


Figure 4.38: **LUR regression coefficients.** Regression standardized coefficients for the land use regression model, LUR.

Predictions comparison

Figure 4.39 depicts the spatial distribution of the average predictions for each of the models for all the study period. In addition, I estimated and mapped the difference between the RF and LUR model predictions (ΔTa, °C) for each pixel. In general, the LUR model captured the same spatial patterns as those observed in the RF model and described above, with some exceptions. For example, for the area nearing the coastline and the northern parts of the prediction area, the LUR model seemed to predict higher temperatures than the RF model (ΔTa$<$ 0 °C); whereas the RF predicted slightly higher temperatures (ΔTa=0-0.5°C) in the central part of the study domain, where the built-up areas are (Figure 4.39(c)). Looking at the distribution of the ΔTa (Figure 4.39(e)), it was apparent that for most of the pixels, the RF model predicted higher temperatures, and that the largest differences were found for pixels where the LUR model predicted higher temperatures (left skewed histogram). The scatter plot (Figure 4.39(d)) shows that the largest differences occurred in the high end of the temperature spectrum, that is, in the warmest temperatures. Finally, there were some pixels which the RF tended to predict as having temperatures around 22°C and which the LUR model gave a much larger range of predictions (between 17 and 20°C). This corresponded to the coastal areas around the peninsula of Santos, which can also be visually seen in Figure 4.39(c) where these areas have a ΔTa=$>$1°C. This was observed across all temporal scales. Further investigations are needed to understand the reason for these differences. For now, care should taken into account if the model is

to be used for studies assessing those areas.

Regarding the temporal variability of the predictions, Figure 4.40 provides a visual comparison of the spatial distribution of the annual average daily mean temperature predicted by the RF and LUR models, and of ΔTa . The overall spatial patterns described for the whole study period persisted more or less across the years, with a negative ΔTa found primarily in coastal and northern areas, and positive ΔTa in the center. However, the level of infiltration of the negative ΔTa over built-up areas varied across years. For example, 2016 showed large areas across the central part with negative ΔTa; whereas in 2017 the negative ΔTa was restricted to the coastal areas.

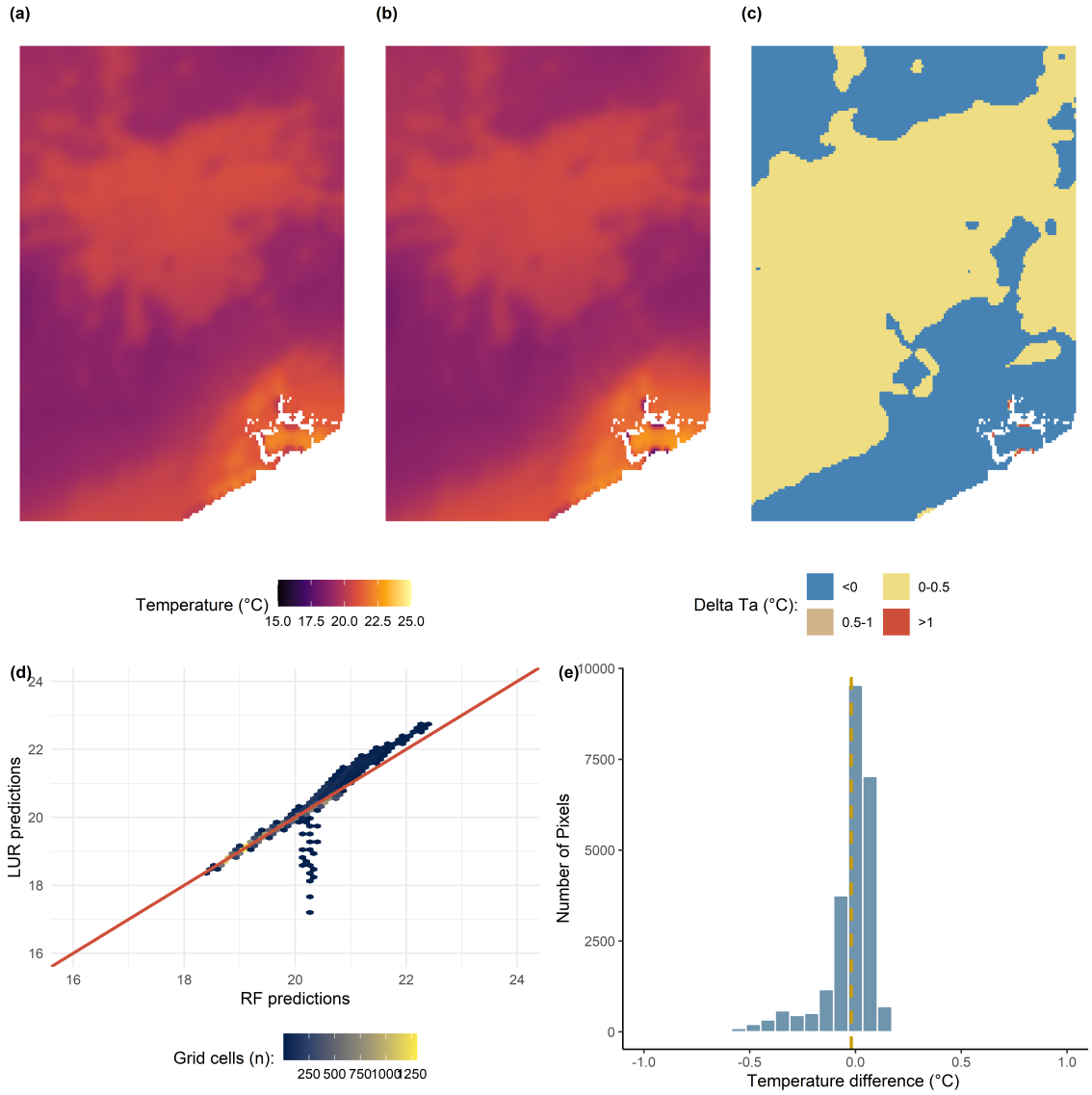


Figure 4.39: **Map of the average daily mean temperature predicted by RF and LUR for all period.** Visual comparison of the spatial distribution of the average daily mean temperature predicted by (a) RF, (b) LUR and (c) the delta temperature between RF-LUR model predictions over the entire study period (2015-2019) (d) Scatter plot between the average daily mean temperature predicted by the RF and the LUR models. The red and blue lines represent the 1:1 line and the linear regression, respectively. The R^2 and regression equation are shown for each plot. (d) Histogram of the distribution of delta temperature values.

In order to facilitate comparison between RF and LUR model predictions over time, I also provided box plots of the LUR and RF predictions overall, by year and by month (Figure 4.41). When looking at the whole period, the differences are negligible (mean$_{LUR}$=19.74°C and mean$_{RF}$=19.72°C). Similarly, the differences between annual estimates are small. Nevertheless, we observe larger differences in the monthly predictions. In short, RF seemed to predict lower temperature for the months of April to August, which happened to be some of the coolest months; whereas, it predicted higher temperatures for the hot months of November, December and January. This is corroborated by the location of the scatter plot points below and above the 1:1 line, respectively in Appendix Figure 6.12.

Performance comparison

Figure 4.30 and Figure 4.42(a) shows the scatter plots between the daily mean temperature predicted by the RF and LUR models, respectively, and that measured by the meteorological stations, assessed through 10-fold spatial-CV. Overall, the LUR model had a slightly poorer performance (RMSE=1.02°C and R^2 =0.92) than the RF model (RMSE=0.80°C and R^2 =0.95) (Table 4.7). The superiority of the RF was maintained across all the groups investigated (i.e., by meteorological station network, year, month and urbanicity). Looking at the accuracy scores for LUR within each explored group, they followed similar patterns as the RF model. For example, the two worst performing meteorological station networks according to the RF model, the *cgs* and *gss noaa* networks (RMSE=0.84°C and RMSE=0.80°C, respectively) , were also the worst performing in the LUR model (RMSE=1.06°C and RMSE=1.05°C). Similarly, the cold months of June to October showed the worst performance (RMSE ranging from 1.08°C to 1.30°C) as opposed to hotter months. Finally, the LUR models also performed worst in rural settings (RMSE=1.15°C and R^2 =0.89) than in urban areas (RMSE=1.00°C and R^2 =0.92). Density scatter plots of the temperatures observed and predicted by the LUR model by year, month and urbanicity can be found in the Appendix (Figure 6.13, Figure 6.14 and Figure 6.16, respectively).

Finally, the RF was also superior when evaluated using the hold-out dataset. In Figure 4.42(b) we can see that despite the fact that the LUR model predicted values were in good agreement with the the hold-out observations (average R^2 =0.90; station-to-station variation: 0.79-0.95), this was poorer than that observed for the RF model (R^2 = 0.92; station-to-station variation: 0.81-0.98, Figure 4.36).

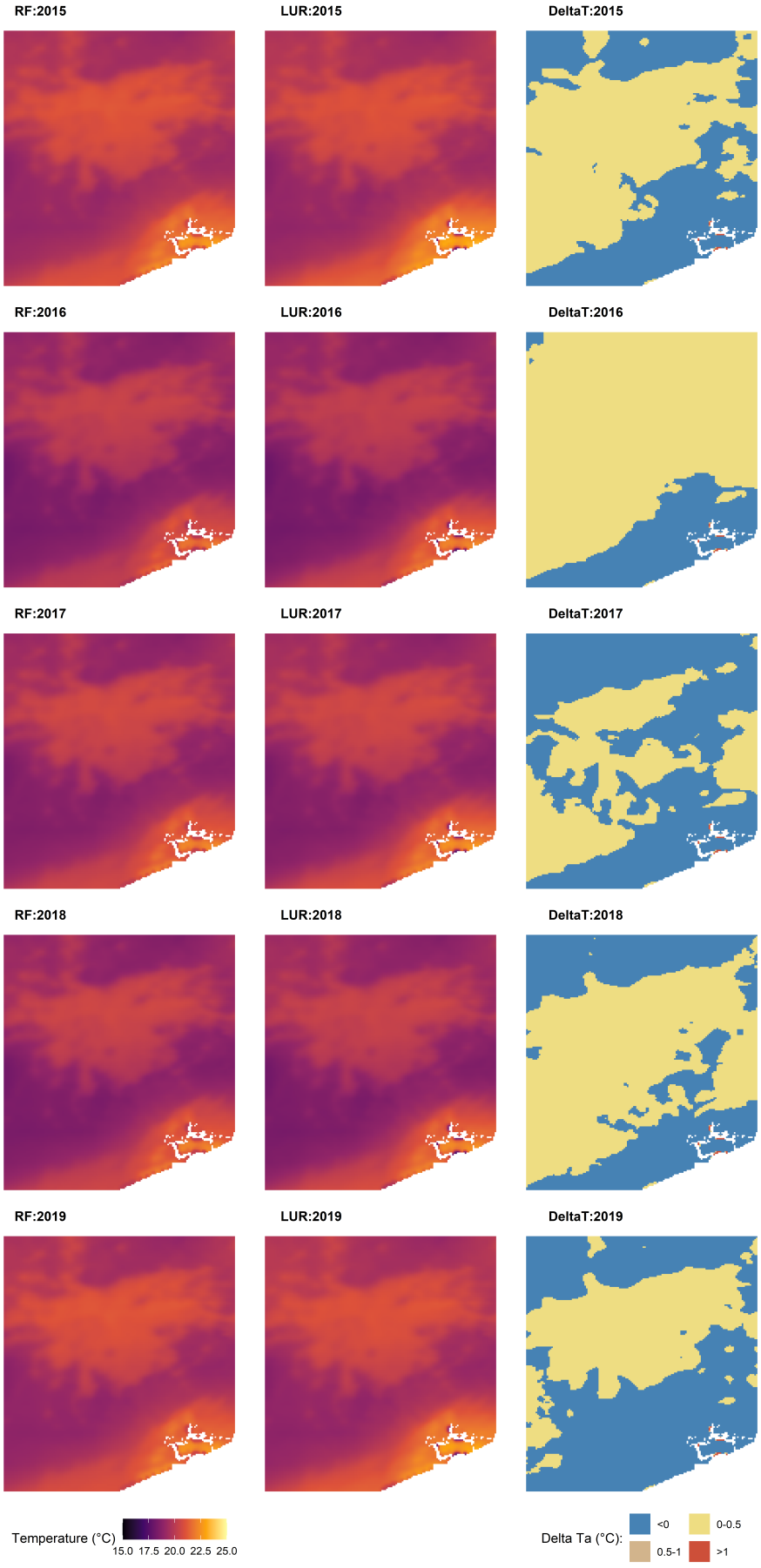


Figure 4.40: **Map of the average daily mean temperature predicted by RF and LUR, and its difference, by year.** Visual comparison of the spatial distribution of the average daily mean temperature predicted by (a) RF, (b) LUR and (c) the difference between RF-LUR (deta Ta) by year.

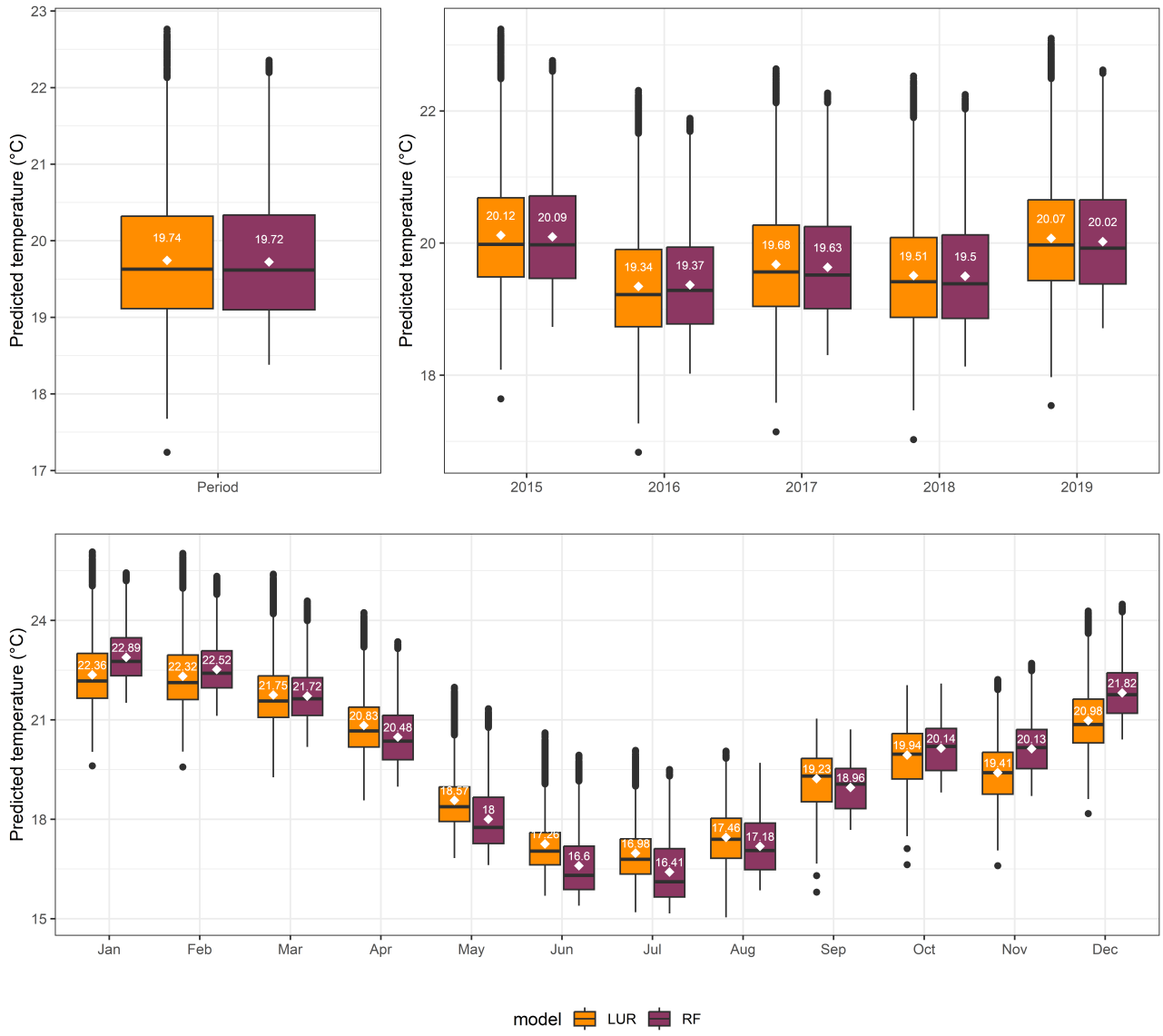
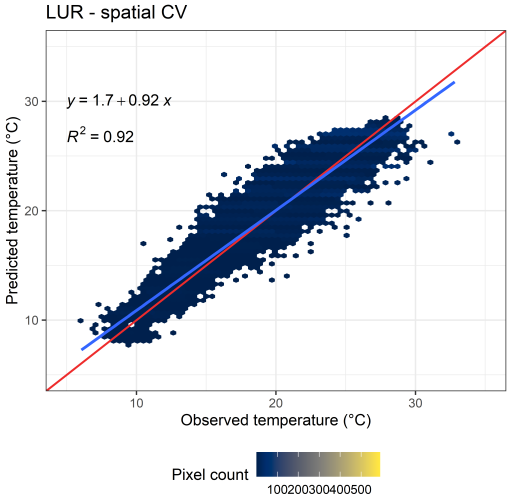
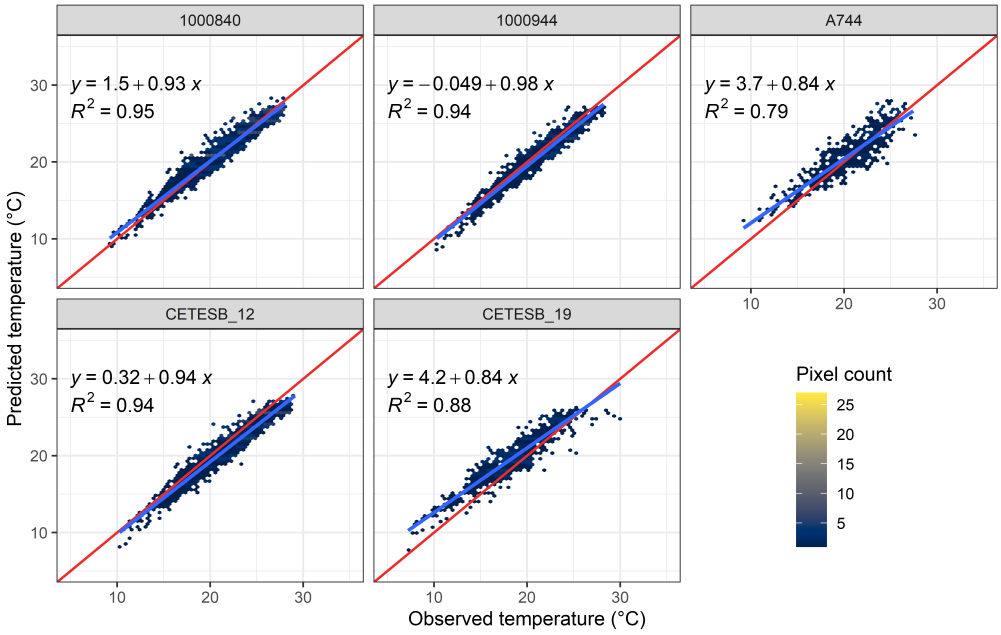


Figure 4.41: **Boxplot of the average temperature predicted by RF and LUR overall, by year and by month.** Boxplot of the daily mean temperature predicted by the RF and LUR models (a) over the entire study period, (b) by year, and (c) by month. Lower and upper box boundaries indicate the 25^{th} and 75^{th} percentiles, respectively, line inside box median, lower and upper error lines $10^{th}$and 90^{th} percentiles, respectively, filled circles data falling outside 10^{th} and 90^{th} percentiles



(a) Scatter plot of LUR spatial CV



(b) Scatter plot of LUR external validation

Figure 4.42: **Scatter plots of LUR spatial CV and external validation.** Density scatter plots of predicted daily mean temperature versus daily mean temperature using (a) 10-fold spatial CV, and for (b) the five hold-out meteorological stations (external validation). Red and blue lines represent the 1:1 line and the linear regression, respectively. The R^2 and regression equation are shown at each plot.

4.3.4 Sensitivity analyses

Table 4.10 show the model performance metrics for each sensitivity analysis. The model performance for the three sensitivity analyses, overall and across all groups, was virtually identical to the main model. When looking at the spatial patterns (Figure 4.43) we also see good agreement across models. The lack of a major improvement in model performance and accuracy by any of the alternative tested variable combinations supports the main model parametrization.

Table 4.10: **Model performance and accuracy by sensitivity analysis.** Performance scores (RMSE in °C, and R^2) for all sensitivity analyses for the whole study periods and by meteorological station network, year, month and urbanicity, assessed through spatial-CV.

	Main model		SA_coast		SA_d2m		SA_both	
Group	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
Total	0.803	0.947	0.803	0.947	0.802	0.947	0.802	0.947
airport	0.738	0.960	0.738	0.960	0.739	0.960	0.739	0.960
cetesb	0.728	0.955	0.729	0.955	0.729	0.955	0.728	0.955
cgs	0.837	0.944	0.837	0.944	0.836	0.944	0.836	0.944
GHCN noaa	0.432	0.985	0.433	0.985	0.432	0.985	0.434	0.985
gss noaa	0.798	0.951	0.798	0.951	0.797	0.951	0.797	0.951
INMET	0.762	0.950	0.762	0.950	0.761	0.950	0.761	0.950
USP station	0.626	0.968	0.627	0.967	0.628	0.967	0.628	0.967
2015	0.798	0.939	0.798	0.939	0.798	0.939	0.797	0.939
2016	0.788	0.960	0.788	0.960	0.788	0.960	0.788	0.960
2017	0.779	0.945	0.780	0.945	0.779	0.945	0.780	0.945
2018	0.837	0.935	0.837	0.935	0.836	0.935	0.836	0.935
2019	0.807	0.950	0.807	0.950	0.806	0.950	0.806	0.950
Jan	0.740	0.893	0.739	0.893	0.739	0.893	0.738	0.893
Feb	0.741	0.888	0.740	0.888	0.740	0.888	0.741	0.888
Mar	0.694	0.863	0.693	0.863	0.693	0.863	0.692	0.863
Apr	0.694	0.929	0.695	0.929	0.695	0.929	0.694	0.929
May	0.749	0.890	0.750	0.889	0.749	0.890	0.750	0.889
Jun	0.843	0.908	0.843	0.908	0.843	0.908	0.842	0.908
Jul	0.936	0.899	0.937	0.899	0.935	0.899	0.937	0.899
Aug	0.856	0.922	0.856	0.922	0.856	0.922	0.857	0.922
Sep	0.944	0.930	0.945	0.929	0.944	0.930	0.944	0.930
Oct	0.855	0.922	0.856	0.922	0.855	0.922	0.855	0.922
Nov	0.764	0.911	0.765	0.911	0.764	0.911	0.765	0.911
Dec	0.750	0.901	0.750	0.902	0.748	0.902	0.749	0.902
Rural	1.053	0.909	1.053	0.909	1.051	0.909	1.052	0.909
Urban	0.765	0.952	0.765	0.952	0.765	0.952	0.765	0.952

RMSE and R^2 values with three decimal places to see differences across groups.

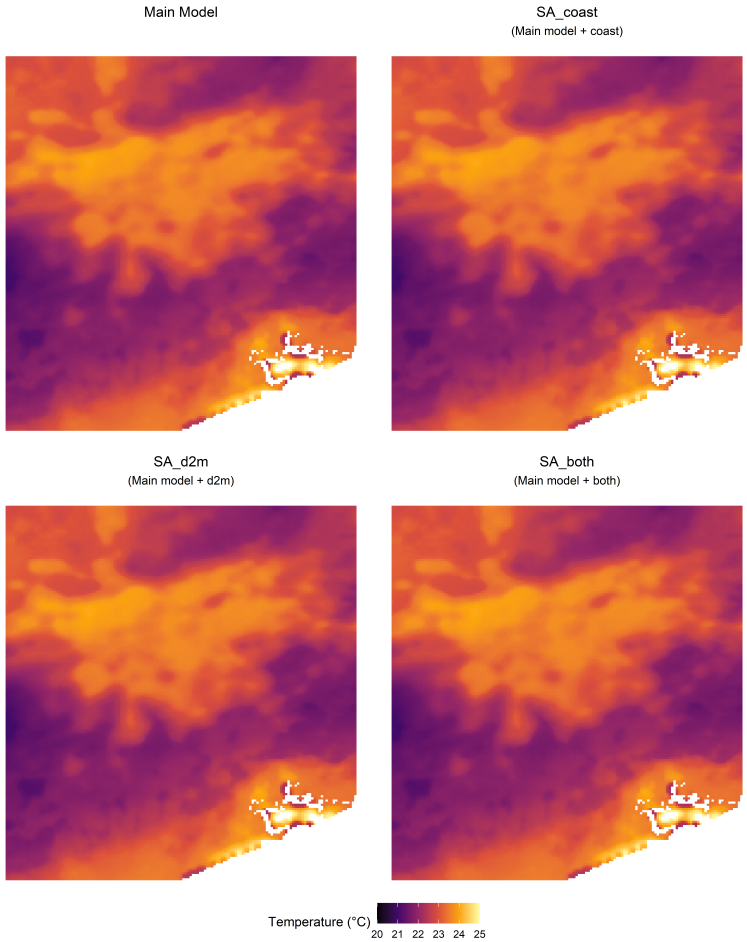


Figure 4.43: **Map of predicted temperature overall, by sensitivity analyses.** Map of the daily mean temperature averaged across all study period as predicted by each sensitivity analyses.

4.4 Discussion

This study showcases the development and implementation of a RF algorithm to reconstruct daily mean temperature in the municipality of São Paulo between 2015 and 2019 at a 500m^2 resolution using multi-source and multi-scale earth observation products and re-analysis data. I evaluated its performance using a spatial-CV which has been suggested to provide more reliable error estimates for space-time models, and a held-out independent validation sample[286]. To the best of my knowledge, this constitutes the first attempt to generate a temperature data product at such high spatial and temporal resolution based on a data-driven approach for São Paulo, and more broadly for South America. The model shows a good performance to estimate both the temporal and spatial variability of ambient temperature across the study area, suggesting its suitability for epidemiological studies. Nevertheless, we observed that the performance of the model was weaker in rural areas, particularly in the south. This dataset has potential applications in a variety of settings such as public health, urban planning, and climatology, among others.

Spatial and temporal variation of predicted temperature

The spatial and temporal variation of the modelled temperature showed that temperature was generally higher in built-up areas, close to the coast and during the months of December, January and February. Below, I describe some of the drivers behind these patterns. It is important to note that, in building the model, I favoured the most parsimonious model, leaving behind several features that were small contributors. This does not mean that these features are not playing a role in the distribution of temperature. It just means that, for this setting, and considering the temporal and spatial native resolution of the features, they were not adding much information to the model. In other cases, the exclusion was due to the fact that there were other highly-correlated features which were retained in the model, which in practice acted as proxies. In short, the exclusion of a feature from the model did not imply that they had no influence, and thus, there is still value in discussing the predictions in light of their distribution.

LST modulates air temperature of the lower layer of the urban atmosphere, and is a key factor in determining surface radiation and energy exchange, and human comfort in cities [288]. Several studies support this correlation [261, 289–292], including one investigating this relationship within the city of São Paulo [293]. Thus, it is often found to be the most relevant variable in studies modelling ambient temperature [111, 261, 287, 292, 294–296]. Nevertheless, the magnitude of the associations, which is influenced by different terrestrial (e.g., refletance, vegetation...) and atmospheric conditions (e.g., radiation, wind, humidity...), can vary substantially. The way by which the air and the surface respond to these external forces is different, which leads to spatial and temporal variations in the association

between LST and ambient temperature, often resulting in a non-linear relationship. Several studies have found substantial heterogeneity in this relationship by ecosystems and regions [114], as well as by season [289] and time of the day [291, 296, 297]. For example, Zhu et al., 2013 [296] found that the nighttime LST, was a good predictor of minimum ambient temperature (RMSE: 2.97, R^2 =0.94); whilst LST daytime had a much poorer performance in predicting maximum temperature (RMSE=7.45, R^2 =0.83). During the day hours more complex and non-linear associations are observed, likely to be caused by interactions with other factors, e.g., satellite–sun geometry, surface geometry, sky view factor, thermal properties of the underlined surfaces, among others. This may partially explain why LST is not the most important feature in my model.

Another reason, which is likely to be more important, is the competing presence of ERA5-Land temperature data products (i.e., *d2m* and *t2m*) in my model. ERA5-Land data does not use observations directly, yet it uses ERA5 atmospheric variables, which are a re-analyses of model and historical *in-situ* variables, to define the atmospheric forcing constrain of the model. This will capture much of the spatio-temporal distribution of the ambient temperature. Moreover, the inclusion on *in-situ* observations, even if second-hand, is likely to capture some of the more specific patterns that LST may not. Thus, several studies suggested the ERA5-Land temperature products for epidemiological studies [36, 79, 298]. Nevertheless, when compared with local monitoring stations, some disagreement was noted across regions and for extreme temperatures. Misty et al., 2022 [36], compared temperature recordings for 612 cities worldwide against the ERA5-Land estimates. They found a global median Pearson correlation of 0.987. However, authors noted a decrease in the correlations in LMIC, in particular in the Philippines, Central America (Costa Rica, Panama), Ecuador, Colombia, and parts of northern Brazil. São Paulo had a 0.95 correlation coefficient (RMSE for station - ERA5-Land=1.03°C). Finally, ERA5-Land has been shown to underestimates temperature, which is more apparent at high temperatures [298] due to relying only on background meteorological stations. This excludes stations in built-up areas, and so, does not capture some of the extreme temperatures recorded in cities where the UHI has a big presence. The lower accuracy of ERA5-Land for high temperatures is of concern as it may lead to important deviations that may affect the identification of extreme events (e.g., heat waves). When producing global health impact estimates, the bias may be negligible, as shown by Misty et al., 2022 [36]. Thus, ERA5-Land has been used previously to assess the acute effects of temperature on health [299]. However, for local studies such us mine, where highly resolved spatio-temporal patterns are of interest, such differences can result in substantial biases. This supports the use of ERA5-Land data in combination to other features, including *lst* to leverage the strengths and limitations of each data source.

From my results, it is also clear that the model picked up on the cooling effect of greenspaces. In the city center, this resulted in rare 'freshness or cool islands' coinciding with the location of some of the

major parks in the city, namely *Carmo Park – Olavo Egydio Setúbal*, *Parque Ecológico do Tietê* and the *Parque de Ciência e Tecnologia da Universidade de São Paulo*. Larger greenspaces in the north and south were also clearly captured by the model, including the *Parque Estadual Cantareira* located in the North of the city, and the *Serra do Mar*, which expands over the South East, following the coastline. NDVI is the primary input feature used to capture green spaces. Theoretically, NDVI influences the temporal and spatial distribution of temperature through vegetation cover which influences the surface thermal conditions [300] and the evaporative control of energy portioning[301], resulting in lower temperatures. After LST, NDVI is probably the most common variable used to model LST in the literature[109, 111, 296, 302–304], with certain approaches developed specifically to be used only with LST and NDVI data, as it is the case of the TVX models [261, 291, 296]. Despite the robust evidence of its influence on temperature, the FFS algorithm deemed *ndvi* as not essential in my modelling framework and thus, was not included in the final model. Upon exploring the *ndvi*- temperature relationship in my study, I found a weak negative association tending to non-linear (Appendix Figure 6.17). This phenomenon has been reported previously in the literature and linked to the presence of additional variables acting as modifiers, such as type of vegetation, latitude, altitudes, climate region and/or season [111, 261, 289, 289, 296]. Moreover, my model included additional variables, such as *lst*, *t2m* or *rh*, which probably captured most of the spatial and temporal variability of temperature influenced by NDVI, making NDVI negligible. Finally, I do not exclude the possibility that the temporal and spatial resolution of the NDVI data may have masked any association. The data was downloaded as 500m^2 resolution averaged over 3 months. During the temporal harmonization data processing, the 3-month estimates were interpolated using linear regression to obtain daily estimates. It is possible that the temporal interpolation has induced some confounding, which is masking the association.

Regarding blue spaces, the effect of the two dams located in the mid and Southern part of the study area (i.e., *Represa Billings* and *Represa de Guarapiranga*) in lowering local temperatures was notorious. As with NDVI, the main features explored that directly captured water bodies, namely lakes and their buffers and the distance to the coastline, were not directly included in the model. In this case, their exclusion was primarily due to their lack of temporal variability (temporally static variables) which caused issues with model over-fitting. Nevertheless, from the model predictions, we can see the other included variables were able to indirectly capture their influence. The spatial distribution of the predicted temperature also suggests that the model was able to capture elements of the topography, such as elevation and slope despite the fact that these were not included in the model implicitly. For example, areas located at higher elevation presented lower temperatures, like it is the case for the area located over the *Cantareira National Park*, which is home to the *Pico do Jaraguá* (1,135 m), the highest mountain in the municipality of São Paulo (Figure 4.27). Their exclusion from

the model was mainly due to their temporally static nature, which cause the *unique-ID phenomenon*, and to the presence of other variables which indirectly captured their influence on temperature, such as *lst* or *rh*, *d2m* and *t2m*.

Finally, a pronounced seasonality was observed in the model outputs (Figure 4.29), which was in line with what was expected for São Paulo. The inclusion of *daylength* aimed at capturing these patterns, reflecting the change in duration of sunlight over a year. Nevertheless, its score in the permutation-based feature importance was rather low (0.023, Figure 4.25) suggesting that most of the temporal variation was being captured by other variables. Clear candidates would be *lst* and *d2m*, which are temperature features and thus, reflect seasonal patterns. Nevertheless, as discussed above, the association between LST and ambient temperature may vary by season [289] and thus, it was important to retain the *daylength* variable to allow the model to detect the seasons.

My model in context

Comparing my model to other available interpolation methods, such as regression and even satellite-based models predicting temperature, is a complex task as there are substantial differences across the studies. These include differences in: target temperature measurement (mean, minimum or maximum temperature); geographic location; size of the area under study, with most studies assessing large areas and only a few focusing on local areas, even less in urban settings; differences in statistical modelling approach; temporal resolution (daily, monthly), temporal coverage (summertime, one year, several years); type, quantity and resolution of predictors used, and density and representativeness of the meteorological stations used to train and validate the models. As mentioned above, there is no other daily average temperature data product available for São Paulo and so, any comparison will be hindered by differences in the factors described above. It is, therefore, difficult to conclude the superiority or inferiority of my model over other models. Even so, there is still value in comparing it to other approaches which I do below.

Compared to other studies estimating specifically daily mean temperature which used different variations of multi-linear regression models, my model (RMSE=0.80°C and R^2 =0.95) showed similar R^2 yet in general, lower model errors (RMSE). For example, Kloog et al., 2014 [305] predicted daily mean temperature over Northeast and Mid-Atlantic USA at a $1km^2$ using a multi-linear regression with day-specific random intercepts and slopes model and obtained a $RMSE_{global}$=2.16°C and R^2_{global}=0.95. In a somewhat complimentary study, Shi et al., 2016 [306] predicted daily mean temperature over the Southeastern USA using a similar statistical approach (i.e., mixed model with day-specific random effects) and input features (i.e., LST, NDVI, elevation, percentage of urban and distance to water bodies). They reported an RMSE of 1.38°C; R^2 of 0.97. Gutierrez-Avila et al., 2021 [304] reported an average RMSE of 1.14 K and a R^2 between 0.78 and 0.95 for the Megapolis of

Mexico City. In France, Kloog et al., 2017 [307] used linear mixed model for predicting daily mean temperature at a 1km^2 spatial resolution, resulting in a global RMSE of 1.68°C and a R^2 of 0.95. My error estimates are closer in magnitude to those reported by Rosenfeld et al., 2017 [308]. In their study, the authors explored the use of MODIS Aqua and Terra LST data products separately for modelling minimum, maximum and mean daily temperature and found that both products produced highly accurate temperature estimates. For mean temperature they reported a RMSE of 0.70 and 0.67°C, and a R^2 of 0.986 and 0.987, for the Aqua and Terra data products, respectively.

Random forests vs other methods

Despite the overall good performance of machine learning models, performance seems to vary substantially by land cover types, year, season or even time of the day [303, 304]. In some occasions, these differences have been attributed to data quality (e.g., missing data missing not at random [303]) or accessibility (e.g., multiple networks of meteorological stations of varying quality standards [304]). However, several studies suggest that the heterogeneity in model performance is likely to be related to the incapacity for the models to capture the complex and non-linear associations that are in place. Thus, several authors have recommended the use of advanced models which allow for nonlinear relationships and which provide a more flexible paradigm to model these varying patterns, such is the case for machine learning algorithms [303]. One of the machine learning algorithms that has received more attention is the RF. In an attempt to confirm this, several authors have compared the performance and accuracy of such models against regression-based linear models or kriging methods [111, 294, 303, 309–311].

Compared to kriging and purely geo-spatial interpolation methods, RF show several advantages. For example they need no rigid statistical assumptions about the distribution and stationary of the target variable. They are also more flexible towards incorporating, combining and extending covariates of different types, which possibly yields more informative maps characterizing the prediction error. Finally, the accuracy and performance of kriging methods relies on the location and density of meteorological stations, suffering from arbitrary and scarce distribution of those [113]. Regression-based methods such as RFs are more independent of spatial density of meteorological stations, and so they give better predictions in areas with low spatial density of available data [312]. Studies comparing the performance of RFs against kriging approaches mostly show comparable or marginally better performance and prediction accuracy [309, 310]. Exceptionally, kriging was shown to perform better and to be less sensitive to bias than the multiple regression methods [313] or than ML algorithms [314]. However, even in those situations, the authors acknowledge that the differences are marginal and that the use of ML algorithms may be justifiable considering that kriging can be computationally demanding, makes many assumptions which are often unlikely to hold, and it may not be easy to

come up with a sound geostatistical model that fits all types of data well [314].

When compared to linear regression approaches, the performance of machine learning algorithms, and in particular of RFs, is superior to the intrinsically linear models. This is supported by my findings. In general, the LUR model exhibited a marginally inferior performance (RMSE = 1.02°C andR^2 = 0.92) compared to the RF model (RMSE = 0.80°C and R^2 = 0.95), in line with other studies. Xu et al., 2014 [311] predicted maximum temperature over British Columbia, Canada, at a 1x1km resolution using a multi-linear regression and RF model. Their results show that the RF model achieved better accuracy (MAE = 2.02°C, R^2 = 0.74) than the MLR (MAE = 2.41°C, R^2 = 0.64). Similarly, Ho et al., 2014 [294] found RF (MAE= 2.31 °C) to outperform linear regression (MAE = 2.46 °C) when predicting daily maximum temperature ($1km^2$) over Great Vancouver, Canada. In a urban study modelling maximum temperature over Greater London, United Kingdom, Dos Santos et al., 2020 [111] compared multiple linear regression against different machine learning algorithms, including RFs, gradient boosting, support vector machine learning and Neural Networks. In their study, Zhang et al 2016 [303] found that complex models accommodating non-linearities (e.g., RF, support vector machine or back propagation neural networks) outperformed intrinsically linear models, particularly when the quality of the input LST data products was poor. The robustness of Random Forest (RF) is likely attributable to an extend to its high adaptability, allowing it to perform well in scenarios with missing values and limited training data. Even when confronted with incomplete input features, RF can provide reasonably accurate predictions. In contrast, Land Use Regression (LUR) models exhibit larger errors and poorer fits in areas with sub-optimal data quality, particularly in rural settings, as evidenced by the higher RMSE of 1.15°C and a lower

RFs have also been compared to other ML algorithms. For example, Mohsenzadeh Karimi et al., 2020 [310] showed the superiority of RF over support vector machines and artificial neural networks for predicting monthly temperature in Iran. In their study, Ho et al., 2014 [294] also compared RF to support vector machines, showing once again ints superiority (MAE of 2.31 °C and 2.46 °C, respectively). However, RF are not always the best performing ML models. In Dos Santos et al., 2020 study [111], the gradient boosting algorithm was the best performing model (RMSE = 2.03 °C and R^2 = 0.68) slightly out-performing RFs (RMSE = 2.13 °C and R^2 = 0.65). Gradient boosting algorithms are a member of the decision-trees algorithms; the main difference with RFs being that the decision trees being built in an additive manner rather than in parallel. In light of the good performance of RF and their computational superiority when compared to kriging methods, they have been suggested for large modelling endeavours. For example, Hashimoto et al., 2019 [315] proposed a NASA Earth Exchange Gridded Daily Meteorology (NEX-GDM) RF model for mapping daily precipitation (among other meteorological variables) at $1km^2$ spatial resolution using satellite, re-analysis, radar, and topography data for the conterminous United States from 1979 to 2017. The superiority of machine learning has

also been reported for the modelling of other environmental exposures, such as air pollution [283] or soil mapping [309].

In short, my study adds to the existing evidence to support the use of ML algorithms, and in particular of RFs, over intrinsically linear and purely geo-spatial methods to model environmental variables. The superiority of RF model is explained by its ability to model complex interactions and non-linear relationships with a relatively low computational and time requirements. By using a combination of hundreds or thousands of trees, RFs are able to efficiently determine which are the strong and weak training trees' settings, allowing more opportunities to find the best model fit, identifying complex associations and providing higher stability [132, 316]. It is worth noting, however, that the superiority of RFs over multi-linear regression is not abysmal and so, depending on the setting, the features available and the computation restrictions, multi-linear regression approach may be a suitable approach. That is, provided that the relationships modeled do not deviate too much from the assumption of linearity. Finally, despite not always superior to other more complex ML algorithms, RFs are attractive as they are really resilient to over-fitting, their results are easier to interpret as opposed to other *black-box* ML models, and the parameter choices has less of an impact on the model performance than for other algorithms [132, 309].

Relevance for urban studies

An adequate exposure assessment is paramount to any epidemiological study, as the opposite, known as *exposure misclassification*, can cause the estimated relative risk of a disease or event to be inaccurate or in the worst case scenario biased. In the case of studies assessing the acute effects (short-term) of temperature, this implies access to accurate daily temperature which is spatially representative of the local environment. The exact spatial resolution will depend on the area being studied; for city-level studies, the desired spatial resolution is of the range of hundreds of meters to capture the micro-climates present in urban areas. Unfortunately, the spatial and temporal resolution of the available temperature data does not satisfy the regional and local needs of epidemiological studies and thus, it is an important research gap [112]. Therefore, the reconstruction of temperature data at high spatial and temporal resolution is paramount from a scientific and practical point of view.

Studies investigating temperature specifically in urban areas are scarce. Some studies have investigated large regions within which there are indeed cities, yet model evaluation has not been conducted locally for specific cities, or at least not reported as such, hindering comparison between urban settings. And only a few studies have investigated urban areas using satellite-based ML models to retrieve Ta [111, 292, 294, 317]. The physical characteristics of urban areas, such as building density and layout, building material or prevalence and distribution of green and blue spaces, significantly modify the energy balance and wind patterns near the surface, making cities particular cases when it comes to

temperature distribution [163–165]. This results in drastic temperature changes within small spatial units (within the scale of just a few hundred meters)[166].

My RF model produced accurate daily mean temperature estimates at a spatial resolution of $500m^2$. This is unprecedented as most studies predicting temperature in urban areas predict at a spatial resolution of $1km^2$ [111, 292, 294, 304]. Breaking this spatial barrier is fundamental for urban settings as temperature can suffer drastic changes within hundreds meters of distance. To my knowledge, there is only one other study producing daily mean temperature using satellite imagery at a higher resolution ($30m^2$) over Tel-Aviv, Israel by Pelta and Chudnovsky A 2016 [317]. In their study, the authors report a performance slightly lower than mine (RMSE= 1.58 and R^2 =0.92, yet comparisons are hindered by differences in the model, area of study, time period and data used. This high spatial resolution was possible thanks to a dense network of meteorological stations within a relatively small spatial extent, i.e. within the city. This is an important advantage over other studies mapping across large regions, such as counties, states or entire countries.

My model was able to identify noticeable difference in temperature between cooler pixels for peri-urban/rural areas and the warmer pixels located at densely built-up areas. The spatial distribution of these hotspots is in line with previous studies reporting on the UHI in São Paulo[24, 196]. This showcases the degree of exposure miss-classification that could be in place if point data was used thus, proving that modelled temperature provides a much better estimate for exposure assessment than point data, i.e., meteorological stations. Nevertheless, meteorological stations are essential in order to train these models and ensure that the predictions are accurate. Thus, there is a need to expand and improve the current network of meteorological stations, specially given the global decline in meteorological stations [21]. Finally, from a global perspective there is a clear bias in the spatial distribution of meteorological station networks towards the Northern Hemisphere, mainly centralized in Europe, China, Japan, Australia and USA [114]. Of all people living in urban areas, about 2.5 billions live more than 25km away from the nearest meteorological station (based on GHCN and GSOD database of meteorological stations which cover most of the stations active in the world)[166]. Thus, there is a pressing need for better coverage of the Southern Hemisphere, particularly in rapidly urbanizing[22] regions in order to monitor spatial and temporal urban extreme heat fluctuations [20, 21, 318].

In addition to its usability for epidemiological studies on the health burden of Ta, a high resolution meteorological dataset like the one here presented can also have major contributions in other fields such as sustainable urban planning [319], hydrological and flooding management [320], building and transport management[319], environmental and ecological studies, or vector-borne disease monitoring [321].

4.4.1 Strengths

The biggest strength of this study, is that it significantly contributes as a compelling proof of concept, showcasing the feasibility of employing satellite images and monitoring stations to generate reliable and accurate temperature predictions at a high spatio-temporal resolution. Although similar endeavors have been undertaken in other regions globally, to the best of my knowledge, this represents the first instance of such an approach being applied in the context of Latin America. The achievement of this milestone unveils promising prospects for the advancement of climate monitoring and analysis in the region, offering invaluable insights for policymakers, scientists, and stakeholders alike. To enable that, upon publication of the research, all data and code will be made open-access and freely accessible on the project's website [26] and my personal GitHub account [25]. I provide a more detailed description of how to access the data and code in the Appendix Section 6.7.1.

The model was built with the parsimonious principle in mind, favouring the model configuration with less features for two main reasons: to avoid over-fitting, and to make our model potentially more transferable to other locations. The method used to select the features, the Forward Feature Selection (FFS), in itself was a remarkable strength of this model. As opposed to traditional selection algorithms which rank features based on importance in the training set, the FFS uses target-oriented CV to evaluate their value for the model. This method was shown to reduce the risk of over-fitting by disfavouring variables with high spatial yet low temporal resolution [286]. Finally, I ensured that all the features used to train the model were open-access and freely available online, making it easy for other researchers to replicate the approach in different settings, update the model or re-train it with new data if needed.

Another important advantage as opposed to other studies, is that I purposely did not include any information about time nor space as predictors, as opposed to previous studies where latitude, longitude or date were included in the model. This not only reduced the risk of over-fitting, as mentioned above, but it also means that, theoretically, the model has the potential to be transferred outside of the temporal and spatial domain used in the training. This is because the model has learn associations between the features and target variable which are not intrinsically unique to the time/space used for training, as it would be the case if the model included latitude, longitude or date variables.

The dense network of meteorological stations, consisting of 43 high-quality stations for a relatively small area, represents another strength of the study. In comparison, other studies typically utilize fewer stations, making this study's network exceptionally robust. For example, Yoo et al., 2018 [292] used only 18 and 23 meteorological stations to train and evaluate a model for maximum temperature over Los Angeles and Seoul, respectively. Similarly, Dos Santos et al., 2020 [111] used 29 to model maximum temperature over the Greater London, UK. This extensive dataset also allows for temperature modeling

at a fine spatial resolution (500m^2), making it particularly suitable for urban studies.

The model validation process further strengthens the study's credibility. The use of a rigorous and spatially explicit evaluation method, 10-fold spatial CV, contributes to my confidence in the model results. Moreover, the availability of a hold-out dataset, possible thanks to the abundance of monitoring stations, enabled a more stringent test of model performance compared to other studies, which lack the sufficient amount of monitoring sites [292, 294], further enhancing the reliability of the findings.

In summary, this study's innovative use of satellite images and monitoring stations to generate high spatiotemporal resolution temperature data with accuracy, along with its rigorous validation process and transparent approach to data sharing, showcases its remarkable strengths and sets the stage for advancements in climate research in Latin America and beyond.

4.4.2 Limitations

While the model exhibits remarkable strengths, certain limitations must be acknowledged.

The first concern is regarding the model's transferability to other settings. The model was specifically designed and trained to capture the temperature micro-conditions in São Paulo, relying on a dense network of meteorological stations and input features which were important for this area. Consequently, its applicability to different regions might be constrained. The model's predictive accuracy is highly dependent on the quality of the input features and the availability of a representative network of meteorological stations. Thus, when applied to other settings and time periods, the accuracy could be significantly affected by the choice of features and ground data used [309, 316].

Secondly, capturing temperature extremes proved to be challenging due to the intrinsic limitations of RF models. From a temporal perspective, this is unlikely to be an issue as our time series covered all time points so we should be able to capture extreme events. From a spatial perspective, extrapolation beyond the training range remains uncertain. The presence of unique temperature profiles without corresponding ground measurements could lead to underestimation of extreme temperatures. However, it should be noted that the majority of meteorological stations are located in built-up areas, covering significant temperature hotspots such as urban heat islands (UHIs). A comparison with a LUR model, capable of extrapolation, indicated that the RF model slightly underestimated extreme temperatures, particularly high temperatures. Nonetheless, given the overall superior performance of the RF model in all other validation tests, it was deemed the most suitable for this modeling task. However, other model-based geostatistical approaches such as spatio-temporal Bayesian kriging can account for non-linearity and complex latent spatio-temporal associations. And so, this issue requires further research both to confirm the extent of this challenge and compare the RFs predictions to a broader range of methodologies, and to investigate more sophisticated techniques such as ensemble models. More

details on the latter in Chapter 6 Section 6.5.1.4.

Another limitation concerns the quality and specifications of input feature data, which significantly impacts RF model performance. Specifically, EO data is subject to issues like cloud and atmospheric contamination, viewing geometry, and sensor degradation, which can vary across products and over time, influencing model accuracy. Moreover, the study considered only a limited selection of open-access features, leaving potential improvement opportunities for model performance and prediction accuracy by exploring other relevant features.

Furthermore, the model's performance and accuracy in rural settings warrant attention. While the model exhibited favorable overall performance with an RMSE of and R^2 between 0.80 and 0.95 °C , the prediction area encompasses a heterogeneous landscape, including urban, peri-urban, and rural regions. The training dataset primarily comprised meteorological stations located in urban settings, with a limited number in peri-urban and rural areas. Consequently, the under-representation of rural areas in the training set might lead to relatively poorer model performance in these regions. Quantifying this difference accurately necessitates abundant ground truth data in rural areas, which remains a challenge. A preliminary analysis comparing model performance in rural and urban settings indicated slightly better results in urban areas. The RF model exhibited a tendency to overestimate recorded values in rural settings, as evidenced by the average absolute difference between observed and predicted temperatures ($\Delta Temp$=-0.22) especially for extreme cold. Addressing this issue requires further investigation and validation using additional ground truth data in rural regions.

In conclusion, while this study has contributed valuable insights and demonstrated the potential of RF models for temperature predictions, the outlined limitations underscore the need for consideration and exploration of potential solutions in future work.

4.5 Summary

This chapter presents one of the first machine learning-based models of daily mean temperature for a Latin American city. The performance of the RF framework was compared to traditional linear regression models, namely Land Use Regression, using spatial 10-fold cross-validation and a hold-out dataset. My results show that, in this context, the RF framework can obtain highly accurate and unbiased predictions, comparable or even outperforming traditional regression methods. Nevertheless, the model validation raised some concerns regarding extreme value predictions and predictions in rural areas, which warrant further investigation. To the best of my knowledge, this is the first ever open-access dataset to provide daily mean temperature estimates at such a high spatial resolution for São Paulo (or any Latin American city), opening new horizons for epidemiological research of the health burden of temperature at an unprecedented granularity in this region.

Chapter 5

Spatio-temporal variability in the temperature-mortality association

5.1 Overview and objectives

While it is well-established that the health burden of temperature is higher in urban areas compared to rural areas, less is known about how the burden varies within a city [13]. The temperature-related health burden might differ by population groups and/or neighbourhood characteristics. The strongest evidence exists for individual-level factors, predominantly showing a strong differential impact by sex and age, and to a lesser degree by other individual-level factors such as SES, occupation, marital status, previous health conditions, ethnicity or education [13]. Evidence on the role of contextual determinants remains scarce, inconsistent and only covers a limited number of factors, such as population density, housing quality, SES, or greenspace [13]. Refer to Chapter 2 Section 2.4 for a detailed overview on the literature around individual and contextual determinants of the temperature-mortality association. Gaining a better understanding of which population groups and/or health conditions are related to an increased vulnerability of the health burden of temperature would help policymakers, stakeholders and urban planners to develop and implement targeted interventions to protect the population against current and future weather risks. This evidence has the potential to enhance and strengthen the response to climate hazards. Findings from high-income countries can seldom be translated to LMIC, due to different climate, and different social, economic and political structures, with often less access to infrastructure, technology and services; all key factors in defining the health burden of temperature [12]. Thus, there is great value in examining local temperature-health risks that can derive specific knowledge to feed into local policies, initiatives and pubic health plans.

Up until recently, most studies investigating the temperature-mortality association have used temperature recordings from one meteorological station, the nearest station or the arithmetic mean across all available stations [40, 112]. Only in recent years, modelled and reanalysis data has received

more attention. Nevertheless, due to restrictions in the spatial resolution of temperature or mortality data, most of the published studies have used cities as their unit of analysis. Kephart et al., 2021 [171] investigated the mortality risk of extreme temperatures in 326 South American cities between 2002 and 2015 using a population-weighted ERA5-Land reanalysis of daily mean temperature. For city-level studies, the use of coarse data, such as that from reanalysis data, seems to compare well to recordings from reference stations, as shown by [36] who compared temperature data from reference stations with ERA5 and ERA5-Land reanalysis data across 612 cities worldwide. For most settings, they found good correlation between the two data sources, and showed negligible differences between the mortality risk estimates obtained from both. However, these comparisons were for city-wide risk estimates, and thus, did not consider the potential exposure misclassification if more granular data is used. Also, they use a reference station per city, which in itself leads to exposure misclassification.

The complex urban layout, wide-spread use of heat-absorbing materials, paucity of greenspaces and the high density of buildings, create complex micro-climates which result in a heterogeneous exposure to temperature. Capturing this is vital in order to accurately assess the exposure in epidemiological studies and reduce exposure bias [253]. In recent years, this challenge has received more attention and a number of studies examining the distinct patterns in temperature-health outcomes within a city have been published [16, 17, 67, 173]. Saucy et al., 2021 [67] examined the effect of temperature on acute cardiovascular mortality in Zurich, Switzerland between 2000-2015. They used $2km^2$ temperature models provided by MeteoSwiss, the national meteorological services. Murage et al., 2020 [17] investigated the role of contextual and individual factors in the temperature-mortality association in London using the MEDMI modelled data from the UK Meteorological Office, available at a resolution of $500m^2$. Bennett et al., 2014 [173] and Gasparrini et al., 2022 [16] used modelled outputs from the UK Meteorological Office at $5km^2$ and $1km^2$, respectively, to explore spatial patterns in the temperature-mortality risk across England and Wales. In all cases, the mortality records contained information of the residential address or postcode, which was used to spatially link to the temperature data. Unfortunately, the availability of temperature data at the spatial and temporal resolution needed for exploring patterns within a city is scarce, often only available for a few of cities primarily in high-income countries [17, 67, 112]. No such data exists for São Paulo to date. This was addressed in Chapter 4 through the development of a Random Forest algorithm to predict daily mean temperature at a $500m^2$ resolution for the municipality of São Paulo between 2015 and 2019.

In this chapter, I examine the relationship between daily mean temperature and mortality for population groups defined by individual and contextual factors in the municipality of São Paulo between 2015 and 2019. I used individual mortality records with geo-referenced residential address and high-resolution daily mean temperature estimates ($500m^2$) in a case-crossover design with a DLNM [59] framework to allow for delayed and non-linear associations.

5.2 Data and Methods

Using geo-coded individual mortality records and high spatially resolved daily mean temperature data, I quantified the risk of all-cause mortality associated temperature using a logistic conditional regression with temperature modelled through a DLNM, stratified by population group. Following the WHO definition of determinants of health [322], I investigated both individual determinants, including sex, age, ethnicity, occupation, education and marital status; and contextual (area-level) determinants, including indicators of SES, housing, employment, occupation, education, and income.

5.2.1 Study design

I used a *time-stratified case-crossover design*. I defined a case as the day when an event (death) occurred. For each case, I defined the controls as the same days of the week within the same month and year. This resulted in a time-stratified case-control match ratio of 1:3 or 1:4, depending on the length of the month and day of the event. This method of case-control matching is known as *asymmetric bidirectional matching*. In detail, controls are selected both before and after the event day with or without restrictions. In my case, the controls were restricted to the same month and year, to account for seasonal and long term trends. This is a specific case of case-crossover time-stratification only possible when the exposure and the event are independent, that is, when the occurrence of an event (i.e., death) does not alter the value of the exposure (i.e., temperature). This design is often used as it provides several advantages: (i) by selecting controls within a short time period around the event date, it ensures that seasonality and longer time trends are controlled for [64, 66]; (ii) it accounts for time-invariant individual-level confounders, such as age and sex, as the cases act as their own controls; (iii) this approach results in a proper conditional logistic likelihood [65]; (iv) it gives unbiased effect sizes and coverage probabilities even with strong seasonal confounding[73]; and (v) as opposed to time-series analyses where temperature is assigned from a unique reference meteorological station, this design allowed me to use precise individual daily temperature estimates; thus, incorporating spatial contrasts in addition to the temporal variability.

5.2.2 Data sources

5.2.2.1 Temperature data

I used the daily mean air temperatures on a 500m^2 grid across the municipality of São Paulo developed in Chapter 4. Figure 4.26 shows the daily mean temperature as predicted by the RF model averaged across the study period (2015-2019) and for a randomly selected day (8^{th} March, 2015) to showcase the spatial resolution of the data. Refer to Chapter 4 for a detailed description of the dataset.

5.2.2.2 Mortality data

In order to examine variability in the temperature-mortality association, it is imperative to have information on the residential address at time of death for each record. *Residential address* is considered personal data, and so is protected and can only be accessed by trained and approved personnel. Mortality data in the São Paulo municipality is managed by the *Coordenação de Epidemiologia e Informação da Secretaria Municipal de Saúde de São Paulo* (CEInfo). Following the CEInfo procedures, and together with my partners at USP, I submitted a geo-coding request for all mortality records between 2015 and 2019. The application required ethics clearance from both USP and CEInfo Ethics Committees. Upon discussion, it was agreed that the CEInfo team would undertake the geo-coding in-house to avoid data breaches. Moreover, in order to reduce the volume of data to be transferred and minimize identifiability, it was agreed that addresses would be grouped by postal code (*Código de Endereçamento Postal*, equivalent to postal code (CEP), an 8-digit code). After residential address, the CEPcodes are the finest geographic unit available, representing a street segment. This is much below the size of the areal unit of analysis of this study, making the loss of spatial resolution negligible. The geo-coding process consisted of the following steps. Firstly, the street name, house number and postal code (i.e. CEPcode) for each death certificate recorded between 2015 and 2019 was collected and checked for accuracy. These checks included spelling mistakes, missing information and/or incoherence. Secondly, the addresses deemed valid were contrasted against the reference database of addresses in the the municipality. Thirdly, the successfully matched addresses where then geo-coded using the "GEOSampa" tool "Address Standardization", developed for this purpose. This resulted in a shapefile of data points representing individual addresses. The rate of successfully geo-coded addresses ranged from 93.97% in 2015 (lowest) to 96.85% in 2019 (highest), with an average of 95.25% for the period 2015-2019 (see 5.1). In order to reduce the volume of the files, individual addresses were grouped by CEPand their CEPcentroid used to represent their geo-location. For validation, a geographic layer of polygons containing the delimitation of postal zones, referring to the 3-digit CEP, was used to identify and exclude points located outside their respective zones. The resulting layer contained the geographic coordinates of the CEPcentroid for which a valid address was identified in the death registry between 2015 and 2019.

The final dataset included information on the date and cause of death, the geographic coordinates of the residential address, and individual characteristics such as age, sex, educational level, marital status, ethnicity and occupation. All data were transferred securely using Imperial College's *File Exchange* platform. This platform uses Secure File Transfer Protocol (SFTP) with shell encryption to provide a high level of security for sending and receiving file transfers. Data were stored safely using the semantic desktop encryption with data access only granted to myself and my supervisor Dr Monica Pirani.I restricted the analyses to non-external deaths (ICD-10 codes: A00-R99) occurring

between 2015 and 2019 among the permanent resident population.

Table 5.1: **Annual success rate of mortality data geo-coding.** Percentage of successfully geo-coded records for the mortality datasets by year.

Year	Total	Successfully geo-coded	
	n	n	%
2015	76,895	72,259	93.97
2016	78,897	74,526	94.46
2017	77,774	74,444	95.72
2018	79,045	75,275	95.23
2019	80,048	77,523	96.85
Total	392,048	374,027	95.25

5.2.2.3 Confounder data

The case-crossover design accounts for time-invariant individual level confounders and potential effect modifiers such as sex or educational level. This is because the cases act as their own controls. However, it does not account for time-varying confounders such can be holidays, air pollution and RH. Here, I considered *holidays* as any national and state-wide holidays. The body's physiological reaction to temperatures is interlinked with RH and so, despite RH in São Paulo displays limited variations on a daily and seasonal basis, I explored including it in the model to verify that it did not confound the relationship.RH was collected from the IAG-USP meteorological station as daily mean RH (%). Similarly, another confounder I considered was air pollution. In the absence of a spatially resolved air pollution dataset, I used the arithmetic mean of PM_{10} estimates across all CETESB-managed [224] ground stations available for the study period and area which records were valid (n=9). Refer to Chapter 3 Section 3.2.2.3 for a detailed description of the data sources and cleaning process for all three variables.

5.2.2.4 Individual determinants

Individual characteristics are those which directly describe the subjects. Here, I used *sex, age group, ethnicity, years of education, marital status* and*occupation*, which I obtained from the death certificates. For sex, I used a binary classification (*females* and *males*) as a more inclusive classification was not available. For age, I focused on 5 year groups starting at *0-14 years old* up to $\geq$ 80 *years old*. Information on ethnicity was provided as per the 2010 census coding: White (census term: *Branco*), Black (*Preto*), Mixed ethnicity (*Pardo*), East Asians (*Amarelo*) and Indigenous (*Indígena*). In order to ensure sufficient statistical power, I combined Black (*Preto*) and Mixed ethnicity (*Pardo*) group, re-coded as *mixed*. East Asians (*Amarelo*) was considered independently, as it has been shown to have different risk profiles, re-coded as *Asian*. Finally, individuals classified as *Branco* were re-coded as *white*. Years of education were re-coded into three categories to ensure contrast, i.e., $\leq$ 3 *years,*

3 to 11 years, and $\geq$ 11 *years*. The original marital status classification distinguished between single, married, widowed, legally separated, divorced, and in a stable union. Social isolation and living alone, are one of the strongest factors suggested to be behind the role of marital status as a modifier of the temperature-mortality association. Based on that premise, and in order to ensure sufficient statistical power, I re-coded them as *single*, which included single and legally separated/divorced groups, *married*, which included those married and in a stable union, and *widows,* which included the widow group. Finally, the occupation field provided in the death certificates was based on the 2002 *Classificação Brasileira de Ocupaçoes* (CBO2002) [323]. It groups occupations in ten groups. I re-coded those in three large groups, i.e., *industry and agriculture*, *commercial*, and *office*, based on the amount of physical exertion often involved and the time spend outdoors (more to less, from left to right). Appendix Table 6.1 provides a detailed description of the original codes and descriptions for all individual determinants and their re-codification. Records with missing information were excluded from the individual determinant specific analyses, which resulted in a different sample size for each individual indicator studied (Table 5.7 and 5.8).

5.2.2.5 Contextual determinants

I defined *contextual determinants* as any indicator or variable which captured a dimension or characteristic of an area. I selected indicators to capture the following contextual dimensions: area-level education, employment, income, SES, housing conditions and demographic characteristics.

I used two main data sources to obtain contextual characteristics: *GeoSampa* and *AtlasBrasil*. The GeoSampa platform is the official consultation system for digital maps in the city of São Paulo (*Sistema de Consulta do Mapa Digital da Cidade de São Paulo*) [277]. The platform contains more than 240 types of information, including on public facilities, transport network, road system, environmental conditions, zoning, historical heritage, and aerial photos, among others. It is the largest collection of geo-spatial data for the city. GeoSampa is maintained by the Municipal Department of Urban Development (SMDU) [324] with the support of several departments, including the Treasury. The Brazilian Human Development Atlas, often referred to as AtlasBrasil, [325] was launched in 2013 with the objective to develop, validate and disseminate a Human Development Index (IDHM) for the 5,565 Brazilian municipalities, 27 Federal Units and for the 20 Metropolitan Regions. It is the result of the joint efforts of the UN Development Program, the Institute for Applied Economic Research (Ipea), and the *João Pinheiro* Foundation[326]. In addition to providing access to the IDHM and its sub-indices, the atlas provides a user-friendly and interactive platform to view and download 200+ indicators on democracy, education, income, labor, housing and vulnerability.

For metropolitan regions, the data is further divided into more granular spatial units, referred to as Human Development Units (HDU). These units are intended to capture the social development con-

ditions at the level of "neighborhoods" (*bairros*), allowing for the investigation of inequalities within cities.HDUs aim to provide a nuanced understanding of intra-metropolitan human development by capturing diverse situations often hidden by aggregate municipal averages. The construction of HDUs involves considering socioeconomic homogeneity and spatial contiguity. The construction of HDUs involves a meticulous consideration of socioeconomic homogeneity and spatial contiguity. Collaborative efforts with research institutions in each metropolitan region are crucial for proposing configurations that accurately represent the region's reality. The HDU construction process utilizes census tract-level socioeconomic information from the censuses of 2000 and 2010 so far. From herein, we will refer to the 2010 information for its temporal proximity to the study period (2015-2019). HDU's are subject to the IBGE confidentiality criteria for spatial units according to which, an area must have a at least 400 permanent private residences. Thus, the areas were constructed following a two-stage process. In the first stage, census tracts are grouped based on homogeneity and continuity, disregarding the 400-residency constraint. In the second stage, HDUs affected by the residency criteria are aggregated into clusters of socioeconomically homogeneous areas, overlooking spatial continuity while favouring proximity. The construction of HDUs respects regional divisions compatible with 2010 demographic census sectors and so, fitting within other territorial divisions, such as districts, as it can be seen in Figure 5.1.

Table 5.2 shows the range of values for some socioeconomic characteristics of the HDUs based on the census 2010. Notably, the income disparity between the highest and lowest income HDUs was substantial, with those in the highest income HDU having a 39-fold greater income compared to their counterparts in the poorest HDU. Moreover, a notable discrepancy of 13.37 years in average life expectancy at birth was observed between the HDUs with the lowest and highest values. Varied educational attainment levels were evident among HDUs, with percentages ranging from as low as 37.9% to as high as 94.15% for individuals above 18 years with at least elementary education. These disparities were effectively captured by the IDHM, which ranged from 0.625 in the poorest HDUs to 0.965 in the wealthiest ones. The wealthiest HDUs, which represented 34% of all HDUs, were mostly located in the city centre and concomitant areas, whereas the poorer neighbourhoods, which represented 26% of HDUs, concurred with peri-urban and rural areas (Figure X). In short, the methodological approach employed for constructing the HDUs ensured spatial homogeneity and contrast across units in terms of socioeconomic indicators.

The selection of indicators to study was guided by the theoretical evidence on contextual vulnerability to temperature-related mortality [149, 159, 179]. Of the relevant indicators, I also took into account the quality of the data and the contrast of that indicator within my cases/controls dataset, to ensure that there was enough variability within my sample to have meaningful comparisons. Table 5.3 summarizes the selected contextual determinants, their source and encoding.

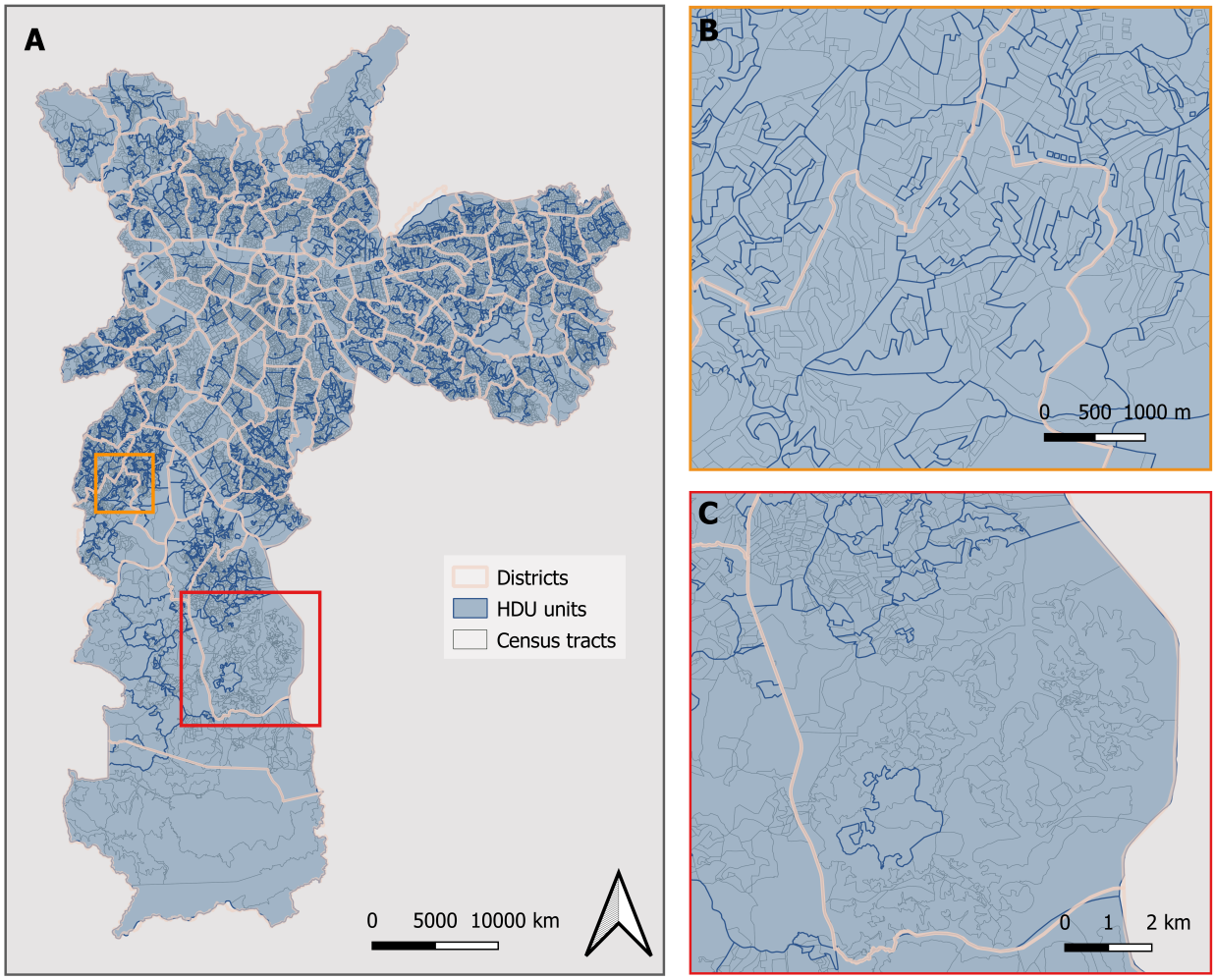


Figure 5.1: **Administrative boundaries São Paulo.** Map displaying the boundaries of the Human Development Units in reference to census tracts and districts for the entire metropolitan area of São Paulo (A), a zoom in of a highly heterogeneous area (B) and of a more homogeneous area (C).

Table 5.2: **Descriptive statistics by HDUs, São Paulo Metropolitan Area.** Highest and lowest values recorded across all HDUs in São Paulo Metropolitan Area for different socioeconomic variables. Based on the 2010 census.

	Lowest	**Highest**
Life expectancy at birth (years)	69.04	82.41
18+ years old graduated from elementary school (%)	37.9	94.15
Average monthly per capita income (R$)	351.85	13,802.96
IDHM	0.965	0.625

Table 5.3: **Contextual determinants selected, codes and description.** Table with all the selected contextual determinants with the general group they belong to, the code, full name, description, spatial unit, year and source.

Group	Code	Full name	Description	Unit	Year	Source
Inequality	idhm_inc	Municipal Human Development Index - income dimension	IDHM sub-index describing income. It is obtained from the indicator income per capita, through the formula. See Eq. 5.1.	HDU	2010	[325]
	idhm_edu	Municipal Human Development Index - education dimension	IDHM sub-index describing education. It is obtained from through the geometric average of the school attendance subindex with a weight of 2/3, and the schooling subindex, with a weight of 1/3. See Eq. 5.1)	HDU	2010	[325]
	idhm_lon	Municipal Human Development Index - longevity dimension	IDHM sub-index describing longevity.It is obtained from the indicator life expectancy at birth. See Eq.5.1.	HDU	2010	[325]
	IPVS2010	Index Paulista of Social Vulnerability	Index Paulista of Social Vulnerability (IPVS) for census 2010	census	2010	[277]
	idhm	Municipal Human Development Index - overall	Municipal Human Development Index. Geometric average of the indexes of the Income, Education, and Longevity dimensions, with equal weights.	HDU	2010	[325]
Housing	water	Water access	Percentage(%) of people living in permanent private dwellings with piped water in at least one room and with exclusive toilet to the total population living in permanent private residences. The water can come from any source (i.e., general network, wells, springs, or from reservoirs supplied by rainwater or water tankers).	HDU	2010	[325]
	crowd	Crowding	Percentage (%) of people living in permanent private dwellings with more than 2 people sleeping in each room, over the total population living in permanent private dwellings	HDU	2010	[325]
	build_mat	Building material	Percentage (%) of households which walls are not made of standard materials (i.e., masonry or finished wood) over the total population residing in permanent private households.	HDU	2010	[325]
Education	illiter	Illiteracy	Percentage(%) of people aged 25 or over who cannot read or write a simple note, over the total number of people in this age group.	HDU	2010	[325]
	educ	Index of educational attainment	Index representing the level of education of the adult population. It is obtained from the indicator "% of 18 years or older with complete fundamental".	HDU	2010	[325]
Demographics	whitePop	Proportion of white population	Proportion of the total population classified as 'white' in the 2010 census	census	2010	[23]
	black_ mixedPop	Proportion of black and mixed population	Proportion of the total population classified as 'black' or 'brown' in the 2010 census	census	2010	[23]
	otherPop	Proportion of other population	Proportion of the total population classified as 'indian' or 'yellow' in the 2010 census.	census	2010	[23]
	elderly	t_env	Percentage(%) of people aged 65 years or older over the total population.	HDU	2010	[325]
Income	gini	GINI index	It measures the degree of inequality in the distribution of individuals according to per capita household income. Its value is 0 when there is no inequality (the per capita household income of all individuals has the same value) and tends to 1 as inequality increases.	HDU	2010	[325]
	empl_inc _ineq	Employment income inequality index	It measures the degree of inequality in the distribution of income of employed individuals aged 18 or over. Its value is 0 when there is no inequality (the income between employed individuals is similar) and tends to infinity when inequality tends to a maximum.	HDU	2010	[325]
	house _pov	Household income poverty	Percentage(%) of people with per capita household income equal to or less than 1/2 minimum wage on the date of the census.	HDU	2010	[325]
	vuln _house	Vulnerable households	Percentage(%) of people living in households vulnerable to poverty (with per capita income below 1/2 the minimum wage in August 2010) and in which most of the income comes from residents aged 65 or over, in reference to the total population resident in permanent private households.	HDU	2010	[325]

Table 5.3: **(Cont.) Contextual determinants selected, codes and description.** Table with all the selected contextual determinants with the general group they belong to, the code, full name, description, spatial unit, year and source.

Group	Code	Full name	Description	Unit	Year	Source
Occupation	outdoor_ manula-boQ	Outdoor manual labour	Percentage (%) of people 18 years old and older employed in sectors primarily outdoors (i.e. the agriculture and livestock sector, the construction sector, and the mineral extraction sector) in reference to the total number of employed people in this age group. Derived from a combination of p_agro, p_constr, and p_extr variables.	HDU	2010	[325]
	indoor_ manula-boQ	Indoor manual moderate labour	Percentage(%) of people 18 years old and older working indoor in jobs requiring strenuous manual labour (i.e. the in the manufacturing industry sector, the trade sector, and the industrial public utilities sector* in reference to the total number of employed people in this age group.Derived from a combination of p_transf and p_siup variables. *This sector includes the activities listed in sections D and E of CNAE DOM 2.0: generation, transmission and distribution of electricity and gas; water collection, treatment and distribution, or sewage.	HDU	2011	[325]
	indoorQ	Indoor labour	Percentage(%) of people 18 years old and older employed in jobs indoors requiring moderate manual labour (i.e. the services sector) in reference to the total number of employed people in this age group. Equivalent to p_serv variable (renamed).	HDU	2012	[325]
Employment	legalemplQ	Legal employment	Percentage(%) of people aged 18 or older without a work contract over the total number of people employed in this age group.	HDU	2010	[325]
	min-wage1Q	Salary below minimum wage	Percentage(%) of people aged 18 or older with monthly income from all jobs the minimum wage, over the total number of employed people in this age group.	HDU	2010	[325]
	min-wage2Q	Salary below twice minimum wage	Percentage(%) of people aged 18 or older with monthly income from all jobs below 2 minimum wages, over the total number of employed people in this age group.	HDU	2010	[325]
	unemplQ	Unemployment	Percentage(%) of people aged 18 or older who were unemployed at the time of the census, over the total number of employed people in this age group.	HDU	2010	[325]
Land use	inf_ settl	Informal settlements	Dicothomous variable identifying ares defined as informal settlements as opposed to the rest. Informal settlements were defined as ZEIS1->5 from the LEI 16.050/14(2014) land use classification.	-	2016	[277]
	rururb	Rural/urban classification	Classification of land into rural or urban	-	2016	[277]
	zonas	mc_ sigla	Code identifying the macroregions as classified by the LEI 16.050/14 (2016) land use classification.	-	2016	[277]

SES is measured through two composite indices, the *Índice Paulista de Vulnerabilidade Social* (IPVS) and the IDHM. A detailed description of each and their constituents sub-indices can be found in Box 5.2.2.5 and 5.2.2.5. Urban planning indicators (i.e. macroareas and informal settlements) are derived from the city of São Paulo Strategic Master Plan which classifies the land based on urban development needs and prospects, providing information on the built environment and socioeconomic conditions of the neighbourhoods. See Box 5.2.2.5 and Table 5.4 for a detailed description. Each contextual determinant was categorized into quintiles. Appendix Figure 6.18 displays the distribution of each variable across all HDUs in the study area (blue) and in my case/control sample (ochre), and the corresponding quintiles. In most cases, the contrast was too small to draw comparison across all groups, and so, I discuss only the extreme groups (1^{st} and 5^{th} quintiles). However, risk estimates for all quintiles can be found in Appendix Figure 6.20 and 6.19.

Box 5.2.2.5 Human Development Index (IDHM) The IDHM is composed of three sub-indices: $IDHM_l$ which represents the longevity, $IDHM_e$ capturing education, and $IDHM_i$ describing income [327].The $IDHM_l$ is defined as the life expectancy at birth. The $IDHM_e$ captures the level of education across the population. It is composed of five indicators; one capturing the proportion of adult population with complete primary schooling, and the other four capturing the proportion of children and youth completing the schooling level appropriate to their age (i.e. 5–6 years old at school, 11–13 years old completing primary school; 15–17 years old completing middle school and 18–20 years old with complete high school). Finally, the $IDHM_i$ is composed of the *per capita* monthly income (in R$ as per August 2010). The index is estimated by calculating the geometric mean of these three components as shown in Eq. 5.1. Values for the IDHM range from 0 to 1 where high values indicate better living conditions and social development. For comparability purposes across time, the Atlas Brasil team often uses a categorical classification into: very high, high, medium, low and very low (see Figure 5.2)

$$IDHM = \sqrt[3]{IDHM_l * IDHM_e * IDHM_i} \tag{5.1}$$

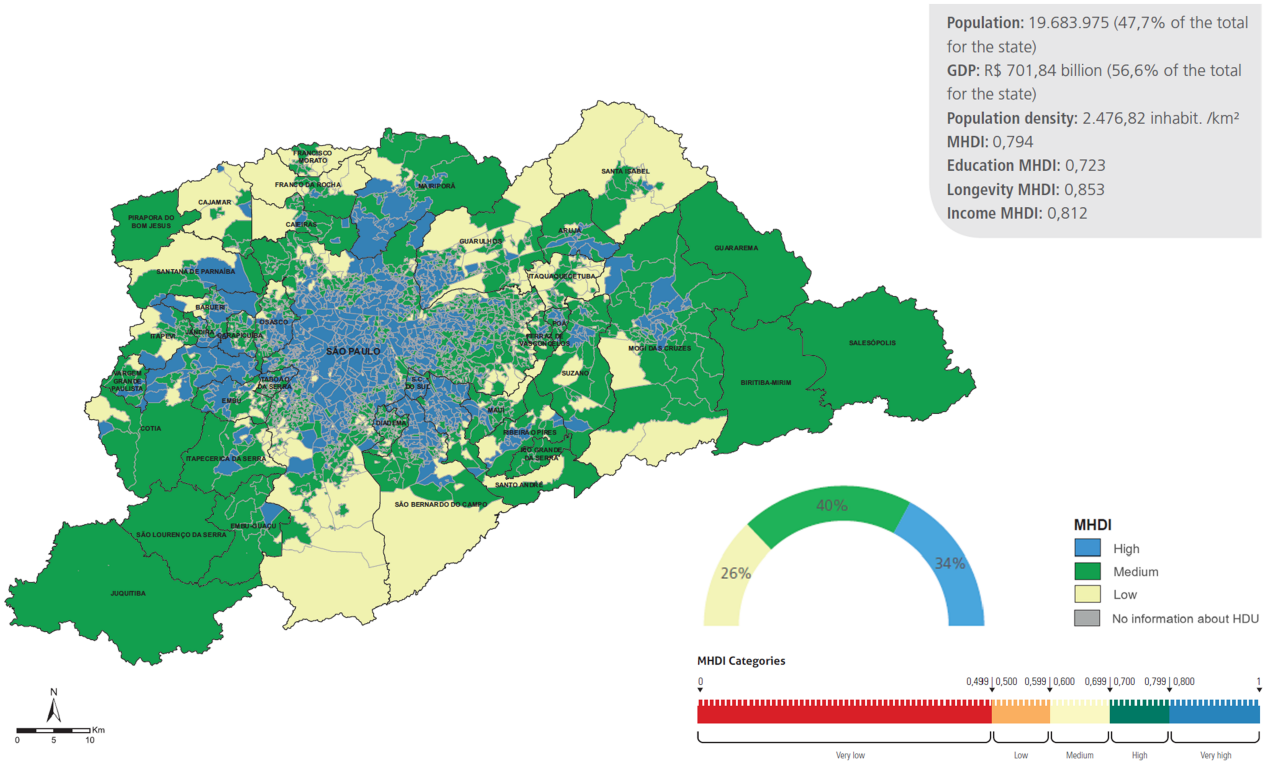


Figure 5.2: **IDHM 2010 in the São Paulo Metropolitan Area** Map displaying Human Development Index (IDHM) categories for the entire metropolitan area of São Paulo. Top right, a panel with some basic statistics; bottom, the proportion of HDUs by category and an illustration of the categorization system employed by Atlas Brasil to categorize IDHM scores. Adapted from [328].

Box 5.2.2.5 Index Paulista of Social Vulnerability (IPVS) The *Índice Paulista de Vulnerabilidade Social* (IPVS) is a composite index of social vulnerability developed specifically for São Paulo. It is composed of five socioeconomic variables (i.e., household income *per capita*, the average income of the woman heads of the household, percentage of households with a household income of up to 1/2 and up to 1/4 of the minimum wage(s) (MW), and percentage of literate household heads) and four demographic variables (i.e., percentage of household heads (overall and females only) between 10 and 29 years old, mean average of age of responsible household heads and percentage of children aged 0 to 5 years old. The variable is combined using factor analysis and cluster analysis, resulting in six groups of vulnerability: (1) lowest vulnerability; (2) very low; (3) low; (4) medium; (5) high, (6) very high (urban), and (7) very high (rural).

Box 5.2.2.5 **Land cover indicators** São Paulo's master plan[213] is a pioneer and new approach to urban planning, putting the addressing of inequalities at its core. It builds on a progressive urban planning agenda that has helped dramatically reduce inequalities in São Paulo over the past 15 years, since the first efforts by Mayor Marta Suplicy in 2002. During that year, the entire city was mapped,including, for the first time, the mapping and census of informal settlements or *favelas*, giving a unique image of the inequalities across the city. This led to the creation of the first *Zonas Especiais de Interesse Social* (ZEIS), which included 138km^2 eligible for the slum-upgrading investment program. Unfortunately, the zoning was mostly unsuccessful as developers avoided investment in such areas and government subsides were insufficient. The new master plan launched in 2014, aimed at addressing these flaws. Among other things, they expanded the zoning with the intention to foster the creation of affordable housing, inject capital in areas of need and to support the transformation of previously unoccupied and run-down buildings and areas into social housing. As part of this, they created a more comprehensive classification of the ZEIS which now includes: ZEIS 1, include primarily informal settlements (or *favelas*) and areas in need of physical upgrading; ZEIS 2 are areas in environmentally sensitive zones; ZEIS 3 include undeveloped peripheral areas; ZEIS 4 are dilapidated areas downtown that need investment, and finally ZEIS 5 are low-density areas that have been deemed optimal to host social housing. ZEIS 1 to 4 were introduced already in 2001, the rest were added in 2014. In addition to this, they also divided the area in eight macroareas based on this susceptibility and urbanity levels and needs. Table 5.4 provides the code used to refer to the macroareas and the full name. Figure 5.3 (A) shows the location of the macroareas and (b) the location of the ZEIS.

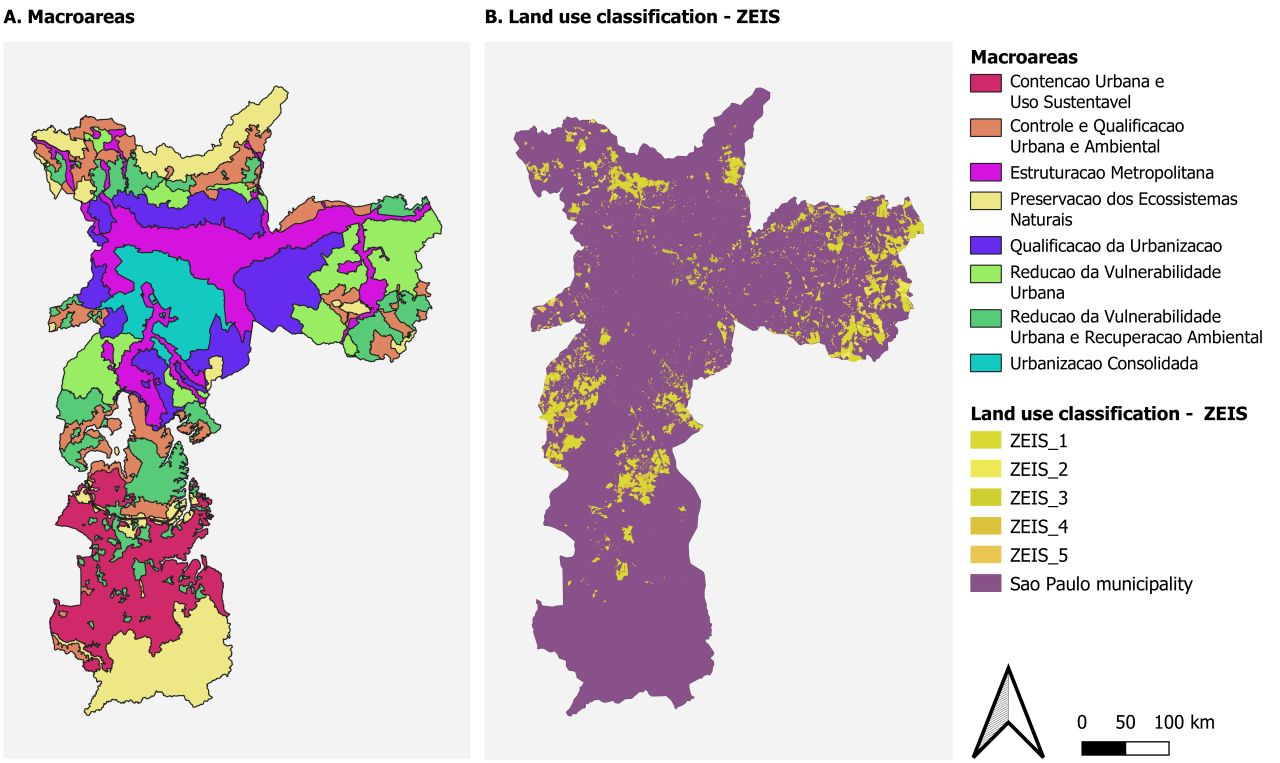


Figure 5.3: **Urban planning variables.** Map of the macroareas (A) and (B) location of the ZEIS areas as defined by the city's urban master plan (LEI 16.050/14 [213])

Table 5.4: **Code and description of the macroareas.** Macroareas as defined by the city's urban master plan (LEI 16.050/14[213]). See Figure 5.3 (A) to visualize the spatial location. Refer to Table 5.4 for the corresponding macroarea codes.

Code	Description
MQU	Macroarea de Qualificacao da Urbanizacao
MCUUS	Macroarea de Contencao Urbana e Uso Sustentavel
MPEN	Macroarea de Preservacao dos Ecossistemas Naturais
MEM	Macroarea de Estruturacao Metropolitana
MRVU	Macroarea de Reducao da Vulnerabilidade Urbana
MCQUA	Macroarea de Controle e Qualificacao Urbana e Ambiental
MUC	Macroarea de Urbanizacao Consolidada
MRVURA	Macroarea de Reducao da Vulnerabilidade Urbana e Recuperacao Ambiental

5.2.3 Data linkage

Figure 5.4 provides an overview of the steps involved in the data linkage of the mortality data with the temperature data and the contextual determinants. All linkages were done based on the georeferenced residential address at time of death. Linkages with temperature data also considered time (date).

As I showed in Chapter 3, temperature can have delayed effects expanding up to several weeks, particularly for cold. Thus, it is important to capture the exposure over a longer period of time. This step requires some additional computation and data manipulation as I needed to reconstruct the

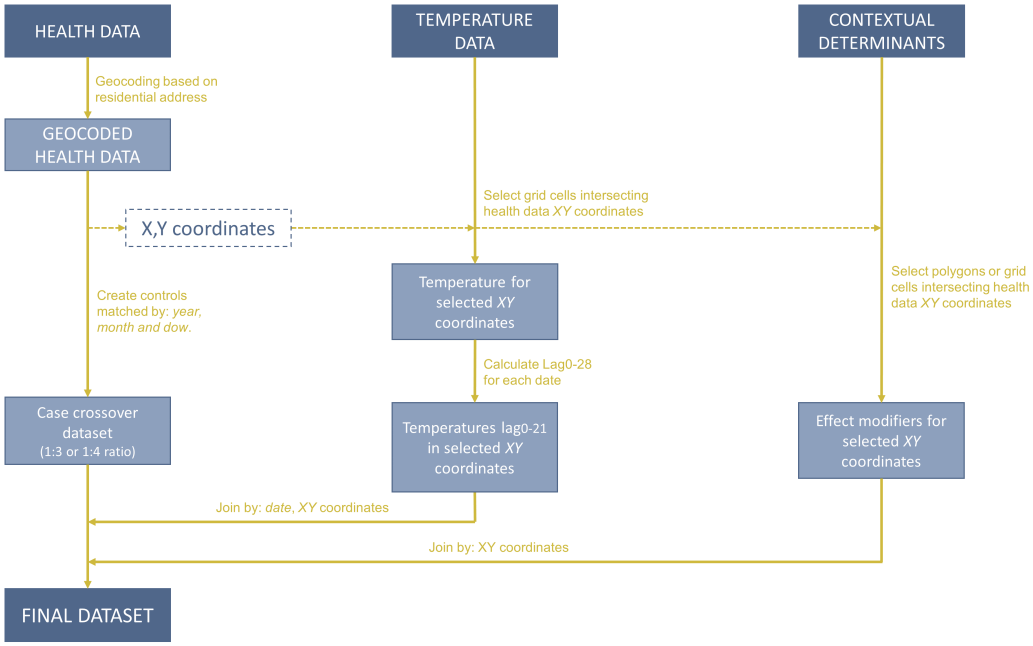


Figure 5.4: **Data linkage overview.** Schematic representation of the steps involved in the linkage of the health, temperature and contextual determinants data

exposure history of each death (both cases and controls) starting from the date of the event backwards to the lag period of interest. To do so, for each death, I extracted not only the daily mean temperature of the case and control dates, but also the daily mean temperature of each day from the date of the event up to 24 days prior. To do so, I created a master grid file of the dimensions and coordinate system of the temperature data (i.e. 500m^2) and assigned a unique ID to each grid, hereafter referred to as *gridID*. I then spatially aligned and intersected the CEP centroid coordinates of all geo-coded addressed from the mortality database to the master grid raster file. Each centroid was linked to a *gridID* according to their spatial location. Using the spatially matched *gridID* and the date, I extracted the daily mean temperature for each case/control dates and the preceding 24 days.

The same approach was followed to link the mortality data with the contextual determinants. In this case, the linkage was done only based on the intersection between CEP centroids and contextual determinants' spatial units, ignoring date as the indicators were time-invariant for the period considered. Finally, confounder data was linked using the date field only.

5.2.4 Statistical Analyses

5.2.4.1 Overview of the modelling framework and model specification

To model the non-linear and delayed association between temperature and mortality, I fitted a conditional logistic regression model with a DLNM to allow for a flexible behaviour of temperature over the lag and exposure dimension [59].

The model fitted to the data was as follows:

$$logit(mort|id) = \alpha + \beta_1 cb_{temp} + \delta PM_{10_{lag0-1}} \quad \text{(Eq.1)}$$

where *mort* denotes the binary variable for the event (case: 1; control: 0), *id* is the subject identifier, α is the intercept and cb_{temp} is the bi-dimensional crossbasis matrix.

The first step was to define the cross-basis functions for the exposure history from the data, and to derive the resulting cross-basis matrix. Here, I defined the exposure dimension as a natural quadratic spline with one knot at the 75^{th} percentile of the temperature distribution, and I used a natural cubic spline with an intercept and three knots equally spaced over the log scale of the temperature distribution for the lag dimension. To define the maximum number of lag days, I tested different options and based the selection on the model performance. Accordingly, I set the maximum lag effect to expand up to 24 days to capture the delayed effects and short-term harvesting [53, 216]. This was the crossbasis function parametrization which resulted in the best model fit (see Table 5.5 for an overview of the crossbasis parameters tested, R1→ R17).

The second step was to consider potential time-varying confounders which are not accounted for by the study design, such as public holidays, RH and air pollution. The *holiday* variable was defined as any national and state-wide holiday day, and was coded as a dichotomous variable. To account for delayed effects in air pollution and RH, I estimated the average of the same day and day before for both. The selection of lag was informed by the literature [56, 215] and my own previous work on temperature-mortality trends [216]. I fitted the model with each of these parameters and combinations of them (Table 5.5, R18→R23). The best performing model, as determined by the lowest AIC value, was the model which included adjustment for average PM_{10} exposure on the day and day before the event [$PM_{10_{lag0-1_t}}$], and thus, it was chosen as the final model (see Eq.1).

As heat and cold-related health effects usually last up to several days and weeks, as shown in my own previous work for São Paulo[216], and to account for potential short-term harvesting, I report the Cumulative Odds Ratio (cOR) over lags 0–24. I estimated the cOR of mortality as the deviation from the MMT to the 95^{th} and 5^{th} percentile of the mean temperature distribution for heat and cold, respectively. For comparability, I used the same MMT (23.29°C) across all population groups, which was derived from the overall empirical MMT for all non-external cause mortality. It corresponds to the 80^{th} percentile of the temperature distribution (overall for total study period and population). Having the same reference point enables me to compare associations between population groups similar to what has been done previously in other studies [67, 172].

5.2.4.2 Stratified analyses by individual and contextual determinants

To investigate whether the impact of non-optimal temperatures on mortality differed by groups of individual and contextual determinants, I modelled the temperature-mortality relationship separately for each category of individual and contextual determinants. This approach allowed me to see possible effect modifications by level of individual and contextual characteristics. Here, I focus on comparing the effects of extreme heat and cold defined as the 95^{th} and 5^{th} percentiles of the overall temperature distribution.

5.2.4.3 Sensitivity analyses

Table 5.5 shows the parameterization for all the sensitivity analyses and the model performance as AIC for each of them. The sensitivity analyses included alternative parametrization of the lag and exposure dimension as well as different combinations of confounders. Among the parametrization tested, I wanted to highlight the exploration of different maximum lags including 21 days which is that found to be the best representation of the risk when using a time-series approach with temperature data from the IAG-USP meteorological station in Chapter 3. According to the AIC, the best lag max for this model was 24 days. When looking at confounder adjustment (R18 $\rightarrow$ 26), the model adjusting for PM_{10} (R20) performed better than that additionally adjusting for holidays and RH (R26). However, previous studies included these variables in their models as confounders. I, therefore, ran a sensitivity analyses additionally adjusting for holidays and RH (R26) for comparison reasons and to ensure that my findings were robust to adding these variables. I also tested a model without confounders (R10). The analysis was run using all non-external causes of death without stratification.

5.2.5 Software

The geo-coding was conducted using R Software Version 4.0.4 and Excel 2016 / Microsoft. For the management of spatial data and visualization of the results, the following software were used: ESRI ArcMap 10.1; QGIS 3.10.6-A Corufia; Padronizador de Enderecos - v 0.1.0.20 from the 'Prefeitura Municipal de São Paulo'; API GoogleMaps Geocoding, and Caliper Maptitude 2012. All statistical tests were two-sided and values of $p<0.05$ were considered statistically significant. All analyses were run in R software (version 4.3.1, R Development Core Team 2009) mostly using the *dlnm* package [77] and the *survival* package [329].

5.2.6 Data access disclaimer

Daily mean temperature at a $500m^2$ spatial resolution was developed as indicated in Chapter 4 and will be soon available for download at my GitHub repository [25] and project's website[26]. Individual

level mortality records were directly provided by theCEInfo as it contained confidential data. To access the data a formal request and ethical approval is required. Air pollution data was obtained from the CETSB website. Hourly RH measurements were obtained upon request from the Institute of Astronomy Geophysics and Atmospheric Sciences and University of São Paulo (IAG-USP). Information on data access in their website. Contextual variables were primarily downloaded from the Geosampa [277] and AtlasBrasil platforms[325].

Table 5.5: **Model parameterization tuning.** Description of all models fitted with their corresponding AIC value. Models R1 → R17 describe the parameter tuning for the crossbasis function of daily temperature; models R18 → R26 describe the fitted models including different combinations of air pollution (PM_{10}), relative humidity (RH) and holidays. Shaded the model selected for it had the lowest AIC value. All models were fitted to all cases/control pairs for non-external causes of death.

Run	**Exposure basis**	**Exposure basis knots**	**Lag basis**	**Lag basis knots**	**LagMax**	**PM10**	**RH**	**Hol**	**AIC**
R1	Quadratic B-spline	1 knot at 75^{th}perc	Cubic B-spline	3 knots+intercept*	21				1017572.615
R2	Cubic B-spline	1 knot at 75^{th}perc	Cubic B-spline	3 knots+intercept*	21				1017579.517
R3	Cubic B-spline	1 knot at 90^{th}perc	Cubic B-spline	3 knots+intercept*	21				1017578.525
R4	Cubic B-spline	2 knot at $50^{th}$and 90^{th}perc	Cubic B-spline	3 knots+intercept*	21				1017587.057
R5	Cubic B-spline	3 knot at 25^{th}, 50^{th}, 90^{th}perc	Cubic B-spline	3 knots+intercept*	21				1017591.319
R6	Cubic B-spline	4 knot at 25^{th}, 50^{th}, 90^{th}, 95^{th}perc	Cubic B-spline	3 knots+intercept*	21				1017595.298
R7	Cubic B-spline	3 knot at 10^{th}, 75^{th}, 90^{th}perc	Cubic B-spline	3 knots+intercept*	21				1017594.567
R8	Cubic B-spline	1 knot at 75^{th}perc	Cubic B-spline	2 knots+intercept*	21				1017581.395
R9	Cubic B-spline	1 knot at 75^{th}perc	Cubic B-spline	3 knots+intercept*	15				1020871.762
R10	Quadratic B-spline	1 knot at 75^{th}perc	Cubic B-spline	3 knots+intercept*	24				1016259.953
R11	Cubic B-spline	1 knot at 75^{th}perc	Cubic B-spline	3 knots+intercept*	24				1016266.626
R12	Cubic B-spline	1 knot at 90^{th}perc	Cubic B-spline	3 knots+intercept*	24				1016265.235
R13	Cubic B-spline	2 knot at $50^{th}$and 90^{th}perc	Cubic B-spline	3 knots+intercept*	24				1016273.716
R14	Cubic B-spline	3 knot at 25^{th}, 50^{th}, 90^{th}perc	Cubic B-spline	3 knots+intercept*	24				1016271.547
R15	Cubic B-spline	4 knot at 25^{th}, 50^{th}, 90^{th}, 95^{th}perc	Cubic B-spline	3 knots+intercept*	24				1016274.977
R16	Cubic B-spline	3 knot at 10^{th}, 75^{th}, 90^{th}perc	Cubic B-spline	3 knots+intercept*	24				1016275.942
R17	Cubic B-spline	1 knot at 75^{th}perc	Cubic B-spline	2 knots+intercept*	24				1016275.348
R18	Quadratic B-spline	1 knot at 75^{th}perc	Cubic B-spline	3 knots+intercept*	24			hol	1016261.728
R19	Quadratic B-spline	1 knot at 75^{th}perc	Cubic B-spline	3 knots+intercept*	24	lag0			1016253.864
R20	Quadratic B-spline	1 knot at 75^{th}perc	Cubic B-spline	3 knots+intercept*	24	lag01			1016252.125
R21	Quadratic B-spline	1 knot at 75^{th}perc	Cubic B-spline	3 knots+intercept*	24		lag0		1016261.356
R22	Quadratic B-spline	1 knot at 75^{th}perc	Cubic B-spline	3 knots+intercept*	24		lag01		1016260.770
R23	Quadratic B-spline	1 knot at 75^{th}perc	Cubic B-spline	3 knots+intercept*	24	lag01		hol	1016253.764
R24	Quadratic B-spline	1 knot at 75^{th}perc	Cubic B-spline	3 knots+intercept*	24		lag01	hol	1016262.584
R25	Quadratic B-spline	1 knot at 75^{th}perc	Cubic B-spline	3 knots+intercept*	24	lag01	lag01		1016253.801
R26	Quadratic B-spline	1 knot at 75^{th}perc	Cubic B-spline	3 knots+intercept*	24	lag01	lag01	hol	1016255.389

Abbreviations: hol, holidays; perc, percentile, and LagMax, maximum lag.
*3 knots equally spaced on the log scale with intercept

5.3 Results

5.3.1 Descriptive statistics

A total of 392,048 deaths were registered between 2015 and 2019 in the municipality of São Paulo. Of these, 374,027 (95.25%) had a valid residential address and could be successfully geo-coded (Table 5.1). I selected only those whose cause of death was registered as non-external, resulting in 350,069 cases included in the study.

Table 5.6 shows the average and standard deviation for cold (5^{th} percentile), heat (95^{th} percentile) and the median (50^{th} percentile) temperature, overall and across all individual and contextual determinants for the entire study period. Appendix Table 6.2 and 6.3 provides the break down statistics for each group. The daily mean temperature ranged from 8.9 to 27.9°C (median: 20.52°C) within the overall study population. For cold temperatures, the mean (SD) across all individual determinants groups was 14.51°C (SD=0.04°C), and across all contextual groups it was 14.47°C (0.18°C), respectively. Similarly, hot temperatures showed little variability, with a value of 25.45°C (SD=0.05°C) and 25.39°C (0.26°C) on average across all individual and contextual determinants, respectively. In general the variability was slightly higher across contextual determinants. This was mainly driven by the notably low cold and heat values recorded for the 'MCUUS' zone with a mean temperature of 13.3°C for cold, and 23.72°C for heat (Appendix Table 6.2). This zone, mainly rural areas in the south of the municipality, with a small population and low mortality count, is known for its comparably lower temperatures (Figure 5.3).

Table 5.6: **Temperature summary statistics.** Summary statistics of daily mean temperature on the day of death overall, and for the dataset of individual and contextual determinants. The *mean* estimates shown represent the average for that statistic across all subgroups in that group (overall, individual or contextual) and the corresponding standard deviation (SD).

	5^{th} perc.		50^{th} perc.		95^{th} perc.	
Group	mean	SD	mean	SD	mean	SD
Overall	14.50	-	20.52	-	25.45	-
Individual	14.51	0.04	20.51	0.05	25.45	0.05
Contextual	14.47	0.18	20.46	0.24	25.39	0.26

Figure 5.5 shows the correlation coefficients for the explored contextual determinants. Most covariates followed the expected correlations. For example, the IDHM was negatively correlated with most employment variables (*minwage1Q, minwage2Q, unemplQ*), *IPVS_V10* and high proportion of black and mixed population composition (*black_mixedpop*), and indicators of standard housing (*crowd* and *build_mat*). Interestingly, it was positively correlated with the *gini* index. This index captures the inequality in income in a defined area looking at the extremes of capital household income. It is possible that the most affluent neighbourhoods contain also some of the largest inequalities.

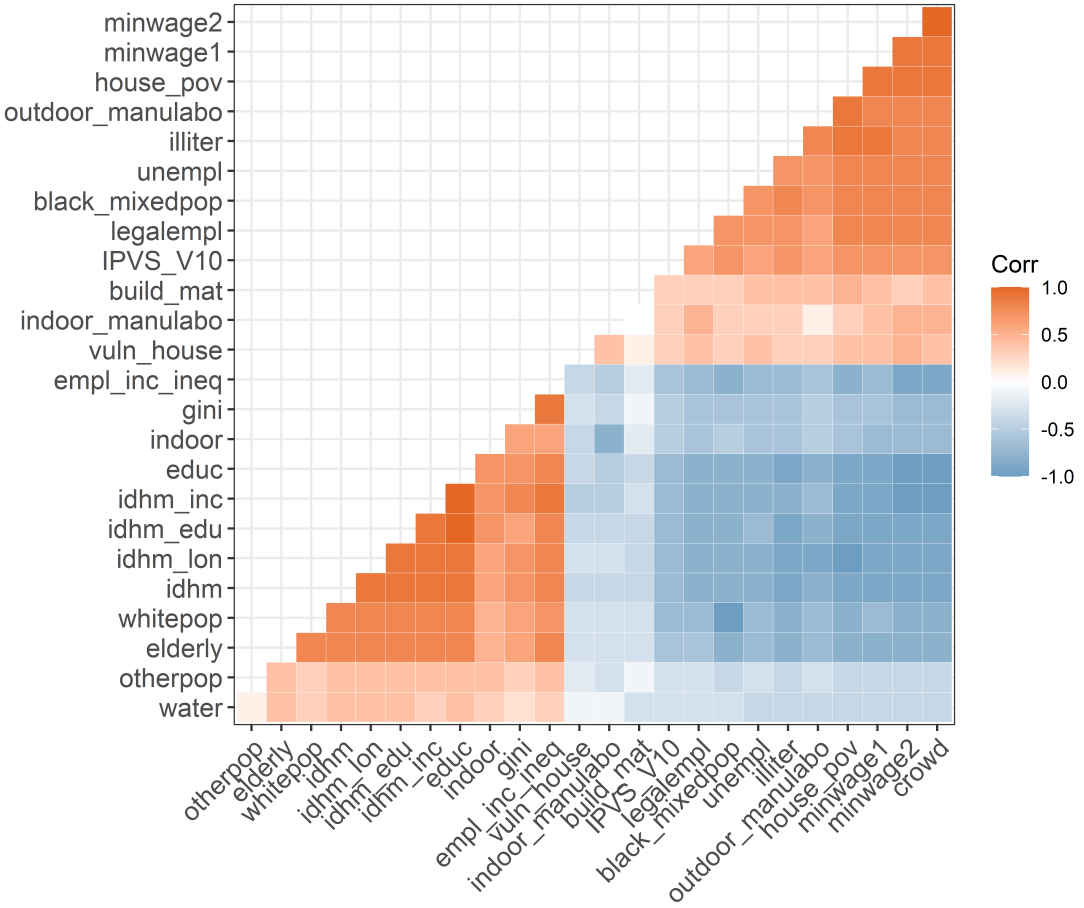


Figure 5.5: **Correlation plot for contextual determinants.** Pearson correlation plot for the continuous version of the contextual variables explored in this analyses. In red, positive correlation, in blue, negative correlation coefficients.

5.3.2 Temperature-mortality association

Figure 5.6 provides different visualizations of the temperature-mortality association for the entire study population between 2015 and 2019. Figure 5.6 (A) shows the three dimensional cOR surface, Figure 5.6 (B) the cOR accumulated over all the lags (24 days) which was U-shaped and resembled that obtained in Chapter 4. Overall, the effects of cold, with a cOR of 1.24 (95%CI: 1.18;1.31), were larger than the effects of heat, with a cOR of 1.05 (95%CI:1.00;1.09) (Figure 5.6 and Table 5.7 first row). The increased cOR for mortality lasted up to 7 days for heat (Figure 5.6 (C)). For cold, the risk increases more slowly, reaching its peak at lag3 to then initiate a paced decrease which lasts up to 24 days (Figure 5.6 (D)). From Figure 5.6 (C) and (D), the presence of harvesting effect is obvious.

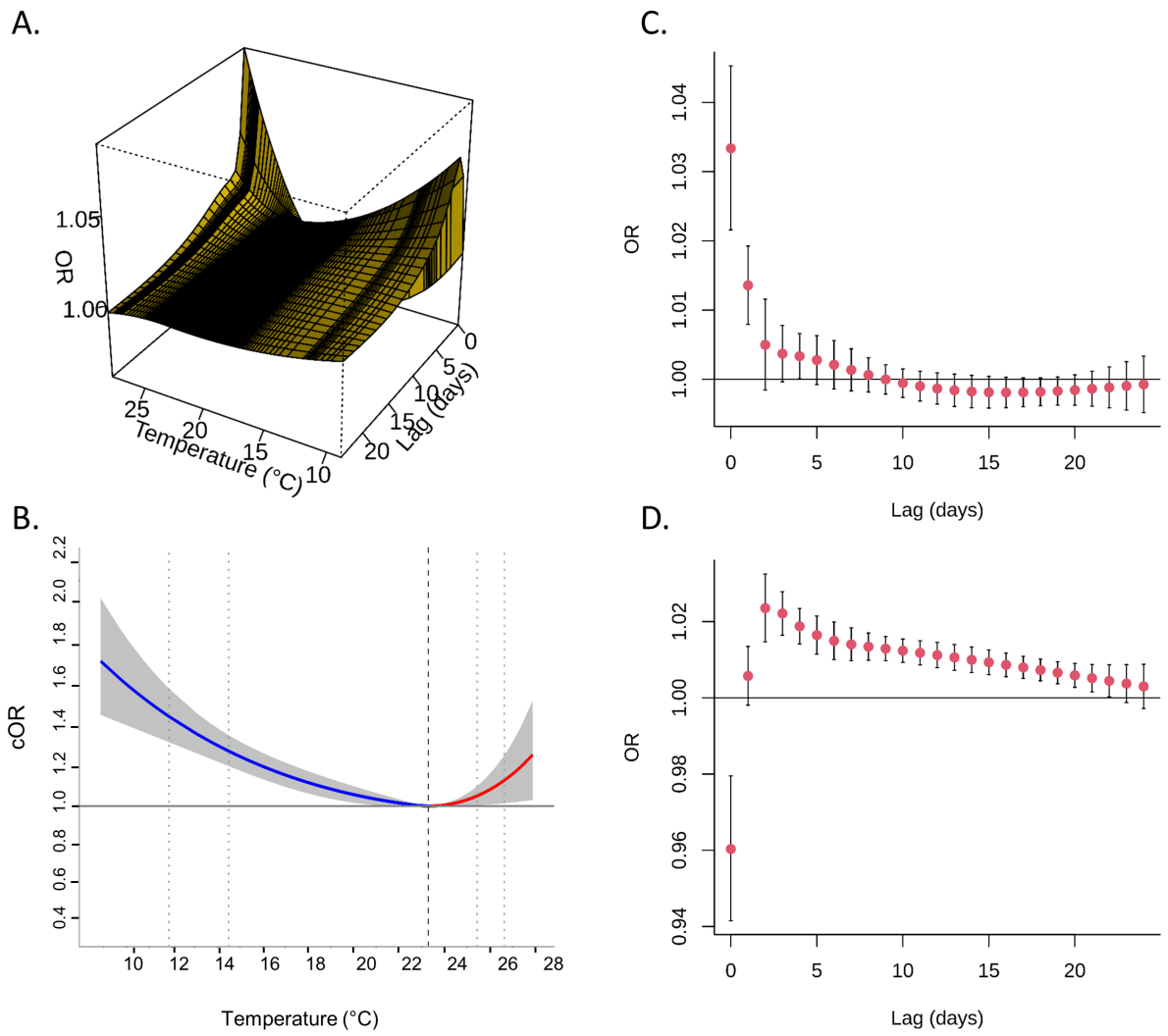


Figure 5.6: **Cumulative Odds Ratio (cOR) of all non-external cause mortality associated to daily mean temperature.** cOR of mortality (A) represented as a three dimensional surface of risk (z-axis) over the lag dimension (x-axis) and the temperature dimension (y-axis); (B) accumulated over all the lags (i.e. 24 days), and (C) specific for each lag between 0 and 24 days for heat (95^{th} percentile of temperature distribution) and (D) for cold (5^{th} percentile of the temperature distribution).

In order to assess whether the results were highly sensitive to model parametrization, I run the overall analyses for the unadjusted model (no confounder variable included, R10 in Table 5.5) and for the fully adjusted model (including holidays, air pollution and RH, R26 in Table 5.5). The results in Figure 5.7 suggest that the main model estimates are robust.

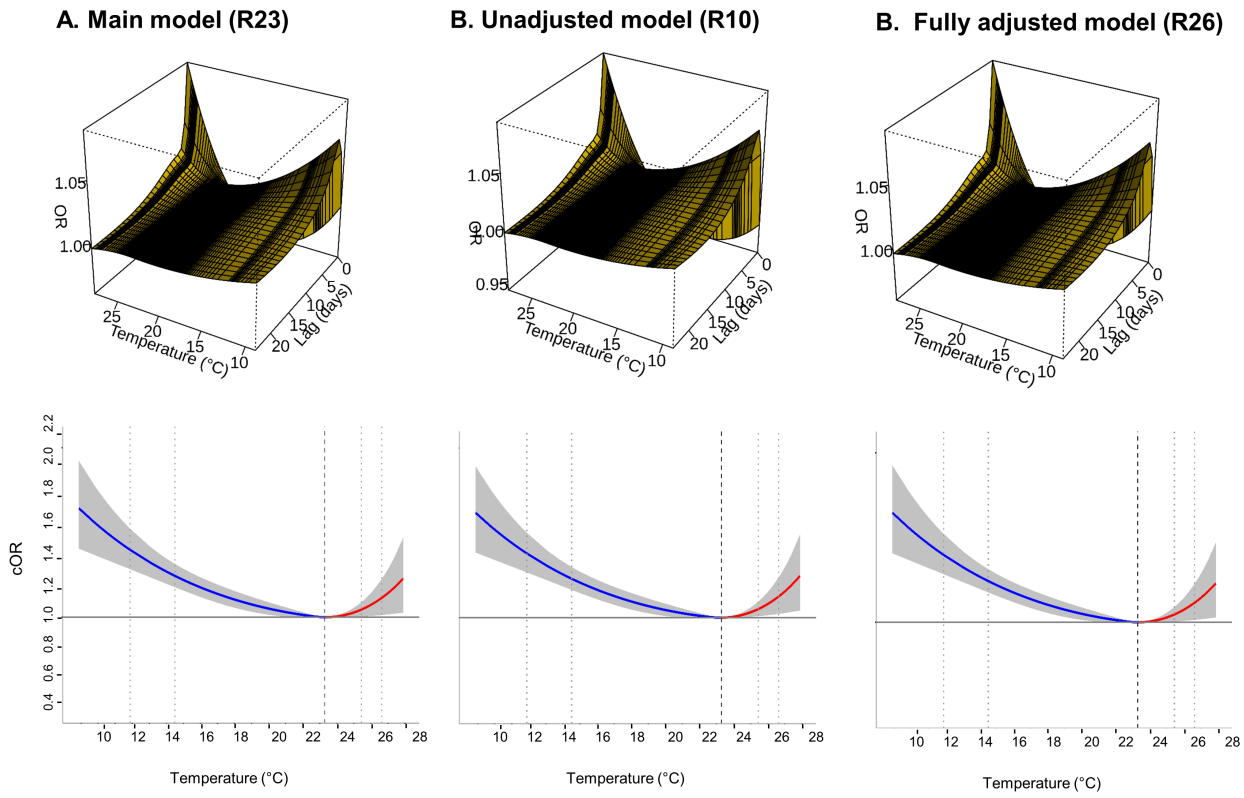


Figure 5.7: **Sensitivity analyses results.**(Top) Three dimensional representation of the Cumulative Odds Ratio (cOR) (z-axis) over the lag (y-axis) and exposure dimension (x-axis), and (bottom) cOR across all 24 days of lag of mortality for all non-external causes of death for (A) the main model, R23*, (B) the unadjusted model, R10*, and (C) the fully adjusted model, R26*. **As per codes in Table 5.5 where the full model parameterization can be found.*

5.3.3 Individual determinants

In general, cold temperatures were significantly associated with mortality across most groups (Table 5.7 and Figure 5.8 (top)). The risk of cold-related mortality increased sustainably with age with the lowest risk registered for the 50-54 years old group (cOR=0.99, 95%CI: 0.77;1.27) and the highest for the 75-79 years old group (cOR=1.44, 95%CI: 1.22;1.70). With respect to ethnic group, the risk was slightly higher among the Mixed group (cOR=1.34, 95%CI: 1.20;1.49) compared to those considered 'white' (1.22, 95%CI: 1.15;1.31). The Asian group showed large uncertainty due to small counts. Compared to those who are single (cOR=1.21, 95%CI: 1.09;1.34), widows (cOR=1.3, 95%CI: 1.17;1.43) and married (cOR=1.29, 95%CI: 1.17;1.41) groups had a slightly higher mortality risk. Across years of education, the risk was the lowest for those with 4-11years of education (cOR=1.21, 95%CI: 1.12;1.31). I also observed increased risk of mortality across all occupational categories investigated, being slightly higher for office jobs compared to commercial jobs. I found no differences between sexes.

With respect to heat-related mortality (Figure 5.8 (bottom) and Table 5.7), most of the associations were not statistically significant with the exception of those aged 80+ years old (OR=1.12, 95%CI: 1.04;1.2), females (OR=1.07, 95%CI: 1.01;1.13) and those with less than 3 years of education (OR=1.16, 95%CI: 1.04;1.24), for whom an increased risk was observed.

Table 5.7: **Cumulative Odds Ratio (cOR) by individual determinant categories.** Cumulative Odds Ratio and 95% confidence intervals (95% CI) for non-external cause mortality for heat and coldv, stratified over individual determinants. Estimates accumulated over 0-24 lag days. Colored, statistically significant results (p<0.05).

Individual determinants	Category	n	cOR $_{cold}$ (95% CI) [a]	cOR $_{heat}$ (95% CI) [b]
Overall	-	1,540,465	1.24 (1.18;1.31)	1.05 (1.00;1.09)
Age group	50-54	16,608	0.99 (0.77;1.27)	1.02 (0.84;1.22)
	55-59	23,160	1.13 (0.91;1.40)	0.88 (0.75;1.03)
	60-64	29,379	1.19 (0.99;1.44)	1.00 (0.87;1.15)
	65-69	33,264	1.18 (0.98;1.41	1.08 (0.95;1.23)
	70-74	34,583	1.34 (1.12;1.59)	1.00 (0.88;1.14)
	75-79	38,949	1.44 (1.22;1.70)	1.06 (0.94;1.20)
	80+	122,456	1.30 (1.19;1.43)	1.12 (1.041.20)
Sex	Females	176,795	1.24 (1.15;1.34)	1.07 (1.01;1.13)
	Males	173,161	1.25 (1.15;1.35)	1.02 (0.97;1.08)
Ethnic group	White	240,066	1.22 (1.15;1.31)	1.04 (0.99;1.10)
	Mixed	89,600	1.34 (1.20;1.49)	1.05 (0.97;1.14)
	Asian	8,139	1.14 (0.80;1.63)	0.97 (0.73;1.29)
Education	<3 years	46,100	1.34 (1.15;1.56)	1.16 (1.04;1.29)
	4-11 years	174,238	1.21 (1.12;1.31)	1.05 (0.99;1.11)
	>11 years	100,300	1.30 (1.18;1.45)	1.03 (0.95;1.11)
Marital status	Single	100,729	1.21 (1.09;1.34)	1.07 (0.99;1.15)
	Married	123,572	1.29 (1.17;1.41)	1.03 (0.96;1.10)
	Widow	103,083	1.30 (1.17;1.43)	1.06 (0.98;1.15)
Occupation	Commercial	31,967	1.24 (1.04;1.49)	0.98 (0.85;1.12)
	Industry and agriculture	45,634	1.27 (1.09;1.48)	1.00 (0.90;1.12)
	Office jobs	50,127	1.30 (1.12;1.50)	1.06 (0.95;1.19)

[a] Cold effects defined as 5^{th} percentile of temperature distribution; [b] Heat effects defined as 95^{th} percentile of temperature distribution.

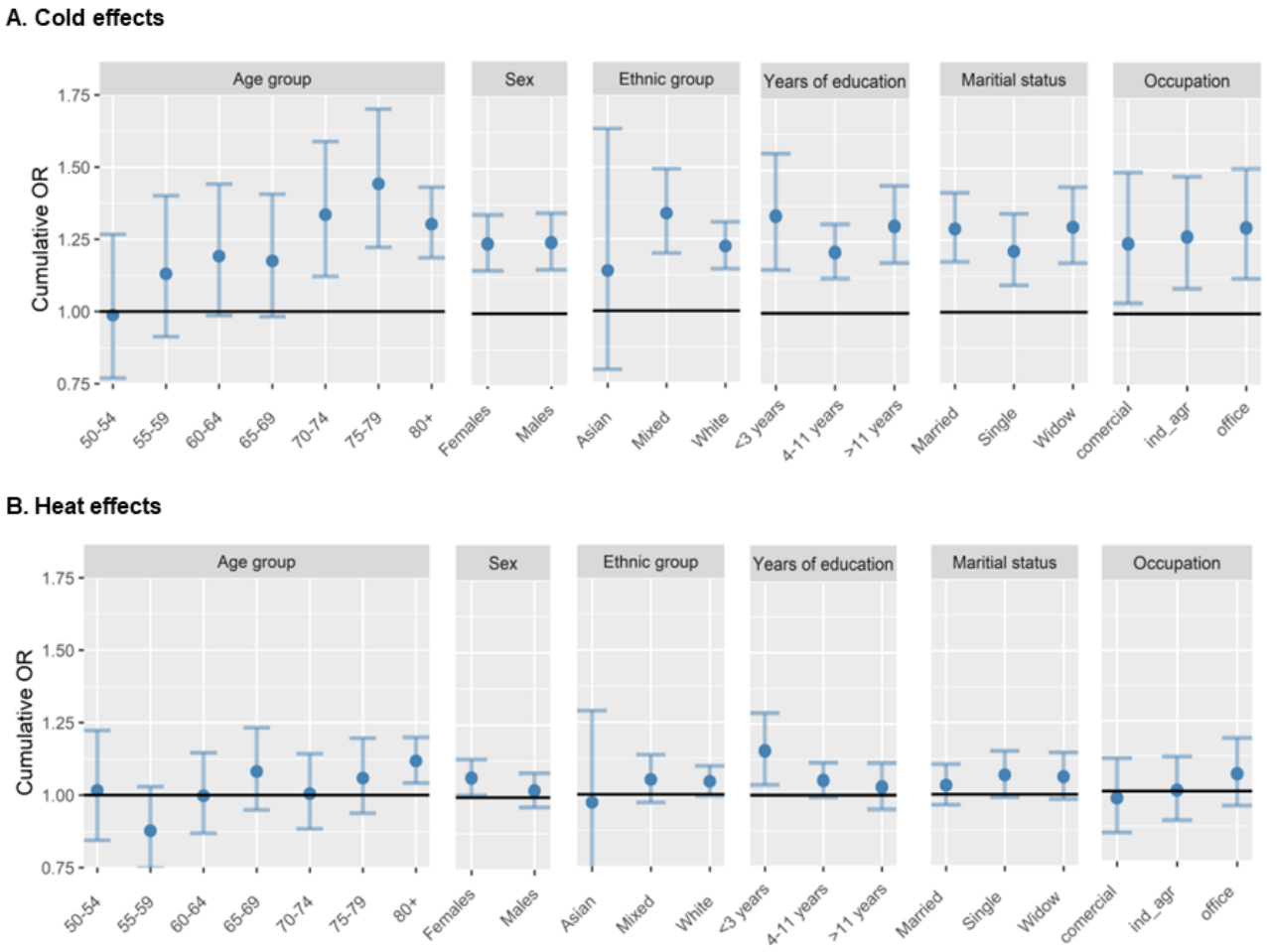


Figure 5.8: **Cumulative Odds Ratio (cOR) by individual determinant categories.** cOR of temperature-related all-cause mortality in relation to the the overall MMT (23.29°C), stratified by age group, sex, ethnic group, years of education, marital status, and occupation (left to right). cOR accumulated over all lags (24 days). The top pane (A) presents cold (5^{th} percentile) and the bottom pane (B) heat-related mortality risk (95^{th} percentile).

5.3.4 Contextual determinants

Table 5.8 presents the cORs and their 95% CIs for the 5^{th} and 95^{th} percentiles of the temperature distribution for each category of contextual determinant. Appendix Figure 6.20 and 6.19 display the risk estimates for all quintiles for cold and heat, respectively. However, in most cases, the contrast of the variables is too small to draw comparisons across all quintiles, as shown in Appendix Figure 6.18. Thus, in Table 5.8, I present only the extreme groups (1^{st} and 5^{th} quintiles). In the reporting of the results below, I draw specific attention to variables for which contrast is small and so, caution should be taken when comparing groups.

Table 5.8: **Cumulative Odds Ratio (cOR) by contextual determinant categories.** cOR and 95% confidence intervals (95% CI) for non-external cause mortality for heat and cold in relation to the the overall MMT (23.29°C), stratified over contextual determinants. Estimates accumulated over 0-24 lag days. Colored, statistically significant results (p<0.05).

			Cold effects[a]			Heat effects[b]		
				95%(CI)			95%(CI)	
Group	Indicator	Class	cOR	lower	upper	cOR	lower	upper
SES indicator	idhm	1	1.25	0.99	1.58	0.9	0.78	1.04
		2	1.09	0.95	1.26	1.04	0.94	1.15
		3	1.28	1.15	1.42	1.05	0.97	1.14
		4	1.28	1.14	1.44	1.12	1.03	1.23
		5	1.26	1.14	1.39	1.01	0.94	1.1
	idhm_edu	1	1.31	1.03	1.66	1	0.86	1.17
		2	1.05	0.91	1.21	0.98	0.89	1.08
		3	1.28	1.15	1.42	1.07	0.99	1.16
		4	1.27	1.13	1.42	1.11	1.02	1.21
		5	1.28	1.16	1.42	1.02	0.94	1.1
	idhm_inc	1	1.16	0.94	1.43	0.92	0.8	1.05
		2	1.15	0.99	1.34	1.12	1.01	1.24
		3	1.26	1.13	1.4	1.05	0.97	1.14
		4	1.29	1.14	1.45	1.07	0.97	1.17
		5	1.27	1.15	1.4	1.03	0.95	1.11
	idhm_lon	1	1.16	0.94	1.43	0.92	0.8	1.05
		2	1.2	1.04	1.4	1.12	1.01	1.24
		3	1.24	1.11	1.38	1.05	0.96	1.13
		4	1.27	1.13	1.43	1.08	0.99	1.19
		5	1.26	1.15	1.4	1.02	0.94	1.1
	IPVS_V10	0	1.04	0.63	1.72	1.11	0.81	1.52
		1	1.2	1.03	1.4	1	0.89	1.13
		2	1.26	1.17	1.36	1.05	0.99	1.12
		3	1.3	1.13	1.5	1.02	0.92	1.14
		4	1.16	0.99	1.36	1.05	0.94	1.18
		5	1.29	1.03	1.62	1.11	0.95	1.28
		6	1.11	0.79	1.57	1.13	0.88	1.45
Housing	build_matQ	1	1.21	1.07	1.36	0.97	0.89	1.05
		5	1.17	1.03	1.32	1.02	0.94	1.12
	crowdQ	1	1.29	1.14	1.46	1	0.9	1.1
		5	1.22	1.08	1.37	1.02	0.94	1.11
	waterQ	1	1.17	1.03	1.32	1	0.92	1.09
		5	1.15	1.02	1.31	1.01	0.92	1.11
Education	educQ	1	1.09	0.97	1.24	0.97	0.89	1.05
		5	1.28	1.13	1.45	0.99	0.9	1.09
	illiterQ	1	1.27	1.12	1.44	1	0.91	1.1
		5	1.22	1.08	1.37	1.04	0.95	1.13
Demographics	black_mixedpop	<10%	1.28	1.11	1.48	1.04	0.93	1.17
		inbetween	1.24	1.16	1.32	1.06	1.01	1.1
		>60%	1.21	0.97	1.5	0.95	0.82	1.1
	whitepop	<20%	1.26	1.03	1.54	0.98	0.85	1.12
		inbetween	1.24	1.17	1.32	1.06	1.01	1.1
		>90%	1.27	1.03	1.57	1.01	0.85	1.19
	otherpop	0%	1.2	0.99	1.45	1.09	0.95	1.25
		inbetween	1.24	1.17	1.31	1.04	1	1.09
		>30%	1.8	1.2	2.69	1.03	0.73	1.44
	elderly	<5%	1.17	1.03	1.33	1.02	0.94	1.11
		inbetween	1.25	1.17	1.34	1.04	0.99	1.09
		>15%	1.3	1.11	1.52	1.13	1	1.28

[a] Cold effects defined as 5^{th} percentile of temperature distribution; [b] Heat effects defined as 95^{th} percentile of temperature distribution.

Table 5.8: **Cumulative Odds Ratio (cOR) by contextual determinant categories.** cOR and 95% confidence intervals (95% CI) for non-external cause mortality for heat and cold in relation to the the overall MMT (23.29°C), stratified over contextual determinants. Estimates accumulated over 0-24 lag days. Colored, statistically significant results (p<0.05).

			Cold effects*			Heat effects*		
				95%(CI)			95%(CI)	
Group	Indicator	Class	cOR	lower	upper	cOR	lower	upper
Income inequality	empl_inc_ineqQ	1	1.2	1.06	1.35	1.03	0.95	1.12
		5	1.31	1.15	1.48	1.04	0.95	1.15
	gini	1	1.24	1.1	1.4	0.96	0.88	1.05
		5	1.2	1.06	1.36	1.03	0.94	1.14
	house_povQ	1	1.24	1.09	1.4	1	0.91	1.1
		5	1.17	1.04	1.32	1.02	0.94	1.11
	vuln_houseQ	1	1.27	1.12	1.44	1.01	0.92	1.11
		5	1.22	1.08	1.38	1.02	0.93	1.12
Occupation	indoor_manulaboQ	1	1.33	1.18	1.51	1.01	0.92	1.11
		5	1.24	1.1	1.41	1.07	0.97	1.18
	indoorQ	1	1.3	1.15	1.47	1.04	0.95	1.14
		5	1.28	1.14	1.45	1.02	0.92	1.12
	outdoor_manulaboQ	1	1.2	1.06	1.36	0.98	0.89	1.08
		5	1.17	1.04	1.33	0.99	0.91	1.08
Employment	legalemplQ	1	1.26	1.11	1.43	1.01	0.92	1.11
		5	1.33	1.18	1.51	1.04	0.95	1.12
	minwage1Q	1	1.27	1.12	1.43	0.99	0.9	1.09
		5	1.2	1.06	1.36	1	0.92	1.08
	minwage2Q	1	1.24	1.09	1.4	0.98	0.89	1.08
		5	1.13	1	1.28	1.02	0.94	1.1
	unemplQ	1	1.26	1.12	1.43	1.09	0.98	1.2
		5	1.23	1.09	1.39	1.04	0.95	1.13
Land cover	inf_settl	formal	1.25	1.18	1.33	1.05	1.01	1.1
		informal	1.21	1.06	1.38	1	0.92	1.1
	rururb	RURAL	1.61	0.91	2.84	0.84	0.62	1.13
		URBANO	1.24	1.17	1.31	1.05	1	1.09
	zonas	MCQUA	1.43	1.08	1.89	1.12	0.93	1.35
		MCUUS	2.28	0.72	7.2	1.08	0.72	1.61
		MEM	1.31	1.14	1.5	1.06	0.96	1.18
		MPEN	3.26	0.44	24.17	0.63	0.21	1.85
		MQU	1.24	1.12	1.37	1.03	0.95	1.11
		MRVU	1.21	1.08	1.35	1.03	0.95	1.13
		MRVURA	1.17	1.01	1.35	1.05	0.96	1.16
		MUC	1.23	1.06	1.44	1.05	0.93	1.19

[a] Cold effects defined as 5^{th} percentile of temperature distribution; [b] Heat effects defined as 95^{th} percentile of temperature distribution.

In general, I did not observe strong indications of effect modification from any of the explored contextual determinants for heat-related mortality, with the majority of groups showing non-statistically significant positive risks. The only groups with statistically significant cOR were those living in formal (1.05; 95%CI:1.01; 1.10) and in urban settings (1.05; 95%CI: 1.0 ;1.09); with more than 15% of elderly population (1.13; 95%CI: 1.00; 1.28); whose ethnic population distribution was between 20-90% of white population (1.06; 95%CI: 1.01;1.10), between 10-60% of black and mixed population (1.06; 95%CI: 1.01;1.10) or between 5-15% of 'other' population (1.04; 95%CI: 1.0;1.09), and those in the fourth quintile of the IDHM index (1.12; 95%CI: 1.03;1.23) and its constituents. The effects by the

proportion of elderly population and by the Gini index were the only ones that showed a statistically positive significant trend, with higher risk with increasing proportion of elderly population and income inequality. Interestingly, the heat-effect for the fifth quintile of Gini index showed an outlier behaviour, which deviated from the general trend (Appendix Figure 6.19). The IPVS index showed a slight increase in risk with increasing deprivation. This contrasted with the increase in risk observed by increasing IDHM development score (i.e., reduction in deprivation). Yet, neither had statistically significant risk estimates. See Table 5.8 and Appendix Figure 6.19 and 6.20.

In contrast, the majority of groups investigated showed a statistically significant for cold-related mortality and a higher risk than that observed for heat. For SES indicators, we observed a significant increase in risk by increasing deprivation as defined by the IPVS index, with a drop in risk in the two most deprived groups. For the IDHM index, similar to heat-effects, higher cold-risks seem to be associated with higher development (i.e., lower deprivation). This was the case also for each of the three IDHM sub-indices. When comparing extreme values of housing determinants, I found statistically significant cold-effects across all groups. Areas with a large percentage of houses made of non-standard materials (*build_matQ*), which are crowded (*crowdQ*) and with piped water supply (*waterQ*) seem to have a smaller cold-related mortality risk, yet differences were marginal. Education determinants seem to have a positive association with cold-related mortality when comparing extreme quintiles, i.e., higher risk for those with higher education. It is important to note that the contrast of the variables is rather small (Figure Appendix 6.18). Areas with a large proportion of households with an income below once or twice the minimum wage, high unemployment rate and low legal employment showed a marginal lower risk than their counter-partners. However, when looking at trends over all quintiles, *minwage2Q* showed a positive trend (Appendix Figure 6.19 and 6.20). It is important to note that except for *minwage2Q*, all variables showed very little contrast (Figure Appendix 6.18). In general, areas with a high proportion of indoor workers either manual (*indoor_manualaboQ*) or not (*indoorQ*) had a lower risk for cold-effects. This was also the case for areas with a large proportion of outdoor manual workers (*outdoor_manulaboQ*), yet the contrast between the fifth and the first quintile was really small in this case (<2% and >7%, see Figure Appendix 6.18). Regarding income inequalities, all contextual determinants explored showed slightly lower cold-related mortality risks for the fifth quintile of their distribution in the study population, except for the employment income inequality. When exploring all the quintiles, no significant trend was observed (Appendix Figure 6.19 and 6.20). Contrast for *vuln_houseQ* and *house_povQ* was rather small (Figure Appendix 6.18). Demographic determinants capturing the proportion of different ethnic groups showed similar risk estimates across groups, except for areas with more than 30% of "other" population (i.e., not white nor black/mixed), yet the uncertainty was really large (1.8; 95%CI: 1.2;2.69). The proportion of elderly people in an area seem to be an important effect modifier, with a positive trend with increasing proportion of

elderly (1.3; 95%CI: 1.11;1.52 versus 1.17[1.03;1.33 for areas with >15% of elderly and <5% of elderly, respectively). Finally, for land cover, in general, I found a marginal higher cold mortality risk for formal (1.25; 95%CI: 1.18;1.33) than informal settlements (1.21; 95%CI: 1.06;1.38). Rural areas had a higher risk than urban areas, yet the uncertainty was too large to draw any conclusions (1.61; 95%CI: 0.91;2.84 vs 1.24; 95%CI: 1.17 ;1.31, respectively). For the urban planning classification, the rural areas of *MPEN* and *MCUUS* showed the highest risks, yet statistically not significant due to large uncertainty. High-build up areas in the city center, namely *MEM*, *MQU* and *MUC*, showed the next highest risks. The lowest risk were recorded for the peri-urban areas of *MQU* and *MRVURA*. See Figure 5.3 for spatial reference.

5.4 Discussion

This study comprehensively explored the impact of temperature on population groups defined by individual and contextual characteristics in the municipality of São Paulo. It is the first study in Latin America to use precise mortality and temperature exposure and to to investigate such a wide range of individual and contextual indicators.In general, the effects of temperature were larger for cold than heat. The strongest effect modification was observed for cold effects by individual determinants with the highest risks among the elderly, those with mixed ethnic background, low level of education, married or widowed and with an office job. For the contextual determinants, I found some weak evidence of effect modification of the cold effect, primarily in relation to land use classification, demographic characteristics and SES indices, with some counter-intuitive findings. Below, I describe my results in the context of the existing literature building on what was already presented in Chapter 2 Section 2.4. Where possible, I focused on studies set in São Paulo, which have been described in that same chapter in Section 2.7 and Table 2.1. Thus, I will refrain from describing their approaches in detail (e.g., study period, statistical approach...) to avoid repetition. I also draw particular attention to studies using spatially resolved temperature data, as this is a critical characteristic of my study. Finally, I provide some potential explanation for the counter-intuitive effects observed for some contextual determinants and discuss some key considerations that must be made when comparing the results.

5.4.1 Overall temperature-mortality association in São Paulo

The shape of the association between temperature and mortality in São Paulo is similar to that described for São Paulo in previous studies [160, 171, 216]. The MMT for the study population between 2015 and 219 was 23.29°C, in line with my previous study [216]. While the heat-related mortality showed a steep increase in short temperature ranges, the cold effect increased steadily over more extended temperature ranges. The delayed effects of cold extended over 24 days, which is

slightly higher than the 21 days reported in my previous study using data from 2000 to 2018 [216]. The observed differences can be attributed to disparities in the study design, specifically the utilization of two distinct approaches: time series and case crossover, and the use of precise temperature exposure as opposed to a sole reference meteorological station. Nevertheless, sensitivity analyses reducing the effects to 21 days showed no significant difference to the findings. Finally, compared to the heat-related mortality, the cold effects were larger in São Paulo. This is in line with previous global studies reporting lower mortality burden for heat in tropical and sub-tropical climate, associated to long-term adaptation to higher temperatures [100]. With the general increase in temperature associated to climate change, cold temperatures will be less frequent. Cold spells are also expected to reduce in the future, as it is the case for the State of São Paulo, as projected by Armani et al., 2022 [197]. Nevertheless, if adaptation to hotter temperatures occurs, as several studies suggest [41, 95, 98] including my own [216], this is likely to result in a higher susceptibility to cold effects thus reducing the overall negative net effect in the total temperature mortality. According to Gasparrini et al., 2017 [198], despite the cold-related excess mortality was projected to decrease across all locations and for the three RCPs investigated, the effects were lesser than those observed for heat, resulting in a net positive change in most locations. This was particularly the case in South America. Thus, cold remains as a hazard worth considering from a public health perspective.

5.4.2 Individual determinants

For cold-related mortality, the highest risk was amongst the eldest group, with trends in increasing risk with older age. In a previous study in São Paulo, Gouveia et al., 2003 [179] reported a slightly higher risk for cold and heat among the elderly group than the adults for the period 1991 to 1994. Later on, Son et al., 2016 [108] would update the findings to 1996–2010 and use more granular age groups. In that occasion, they were able to find a clear trend with increasing age, highest amongst those aged over 75 years old. For heat-related mortality, the trend was present, yet only those aged over 80 years old showed a statistically significant positive risk. The evidence linking age with higher risks of mortality for both heat and cold is vast and consistent, with a positive trend between age and risk [13]. Strong age differences for heat effects were reported by [215] between 1998-2002 and [108] between 1996-2020. These studies used broader age groups (0-15, 16-64 and over 65 years old) which could explain the differences to my results. Finally, Gouveia et al., 2003 [179] reported a slightly higher risk for cold and heat among the elderly group than the adults.

I found a higher risk of heat-related mortality among women. Although, the evidence for effect modification by sex is varied, the majority of studies reported a higher heat effect among women [13]. In São Paulo, this has also been shown to be the case at different time periods: between 1996-2010 [108], between 1985-2005 [203] and between 1998-2002 [215], showing that the gender-gap is

persistent over time. In general, the evidence for cold-effects remain more inconclusive [13]. Of the three studies set in São Paulo, only [108] investigated cold-effects and found slightly higher risk among men. This was not the case in my study, where no statistically significant differences were observed. Several factors have been suggested to contribute to the higher risk observed in women, including a lower sweating ability [330], menopausal effects such as elevated body temperature and sweating [331], behavioral and occupational differences. In a recent review, Steen and colleagues [332] suggested older age as a potential factor contributing to the higher risk among women, while other studies show that the gender-gap remains within age groups [333]. In my study, age was accounted for by design, and so, this is unlikely to explain the differences.

My findings for marital status suggest a higher risk of cold and heat effects among those who are widow, closely followed by those who are married, which is consistent with previous studies [108, 138]. Behavioural differences, social isolation and living alone are often associated with widow groups and have been linked to higher mortality risks for temperature, as well as higher age [17]. Subject-specific time-invariant characteristics, such as sex and age which could be argued to confound the comparability between the single and widow group, are adjusted for in my study by the case crossover design, discarding this as an explanation.

The burden of temperature-related mortality was higher among people with the lowest educational level, particularly for heat. These results are in line with previous studies reporting higher heat-related mortality risk for those with lower or no education [67, 148, 178]. This was also the case for the two studies looking at São Paulo. Bell et al., 2008 [215] described a strong heat effects for those with primary education in São Paulo, which was not seen in other cities. Similarly, Son et al., 2016 [108] found a slight trend of increasing risk with lower educational attainment, with the highest risk amongst those without education. This was not the case for cold-effects, where all groups showed similar effect estimates, except for those with university degrees whose risk was the highest [108]. Studies elsewhere have also reported conflicting results for cold effects [141, 146, 334]. It is difficult to draw conclusions as the variable definition and the importance of the context vary substantially across studies. Education attainment is often considered a proxy for socio-economic deprivation. Yet, the association between socio-economic indicators and educational attainment varies by study area [215].

For occupation, the previous evidence is scarce and heterogeneous. Manual workers have been suggested as a vulnerable group due to their exposure to outdoor temperatures and physical exertion [141, 151]. Considering workers in the commercial sector, in industry and agriculture as manual workers, because their work involves exposure to outdoor temperature and/or physical exertion, compared to office jobs, my findings are not in line with the literature. I found little evidence of a difference between occupation groups; people with office jobs seem to have a slightly higher risk of heat and cold-related mortality. This findings seem counter-intuitive and opposing to the existing literature.

Different occupation classifications and work conditions, and the lack of specificity caused by combining multiple categories to gain statistical power, may account for these differences across studies.

Those from a mixed ethnic background had higher mortality risk than those from a white background, yet the risk was only statistically significant for cold. The lowest risk was observed among people from an Asian background, who showed neither statistically significant risk for heat nor for cold. This is in line with previous studies reporting higher risk of heat effects for blacks and lower vulnerability for Asians [149]. This effects are likely to reflect deeper socio-economic inequalities between the groups [149, 150].

5.4.3 Contextual determinants

In general, the cORs for heat-related mortality across most contextual determinant groups examined were not statistically significant, aligning with the overall lower risk associated with heat that was observed in the general population. Conversely, I found some weak evidence of effect modification for cold-related mortality risk, primarily in relation to land use classification, demographic characteristics and SES indices, yielding some unexpected findings.

Firstly, it is important to consider some theoretical and methodological limitation to socioeconomic area-level indicators. SES is a complex construct which encapsulates multiple dimensions of inequality, such as income, housing conditions, access to services or education. It has been associated with poor health status, restricted healthcare access, and inferior living conditions, such as inadequate housing and limited resources to cope with extreme temperatures [156]. There is no standardised way of characterising SES of a specific area as its most relevant indicator are likely to vary be setting, time period, outcome of interest, and exposure pathways. Income for example, has been shown to be a better proxy for disease progression, whilst education seems to predict better disease onset. Consequently, researchers have investigated the role of individual indicators such as unemployment or literacy rates, or employ composite indices. In this study, both approaches, i.e. assessing SES via a composite measure and specific indicators, were utilized as they provide complimentary information on SES. Regarding individual indicators, it is challenging to find comparable studies, as either the indicators or the thresholds/classification used are often highly specific to the city/country context. Each indicator aims to capture a specific dimension of SES in an area. Here, I used indicators education, occupation, land use, income, demographics employment, housing characteristics, as well as SES indices to capture different dimensions of SES. For each indicator, I also explore potential reasons that may explain the findings. In addition, I provide a more general discussion of potential explanation for the findings obtained for contextual factors in general in the "Counter-intuitive results and comparability considerations" section 5.4.4.

Education. There is suggestive evidence of an increase risk of heat-related mortality risk with

lower education, yet others have found no difference[141, 146, 148, 176]. Results for cold are even more inconsistent [141, 146]. Education could have a direct effect on vulnerability to temperature through lack of awareness or understanding of the risks associated with temperature and/or of the adaptation measures to mitigate them, or through its high correlation with SES [156]. This was not the case in my study, where areas with high proportion of educated individuals had a slightly higher cold-related mortality risk. This contrasts the strong effect modification observed for educational level when looking at individual determinants, and suggests an ecological fallacy issue. Moreover, these two education area-level indicators (*educQ* and *illiterQ*) showed a small contrast in the study population, which is likely to result in noisy results.

Occupation. The prevalence of heightened risk among outdoor manual workers [151] can be attributed to their engagement in physically demanding tasks, primarily conducted outdoors. In the present investigation, regions characterized by a significant proportion of indoor manual workers (*indoor_manulaboQ*) exhibited a diminished susceptibility to cold-related health effects. This observation may be indicative of their reduced exposure to temperature fluctuations inherent in their occupational activities. However, this was not found for the *indoor* group, conflicting this hypothesis. It is crucial to acknowledge that the limited range of variation in occupation types across the studied areas and the lack of precision in their categorization could primarily account for the obtained results. Consequently, it is plausible to assert that the current indicators may not be sufficiently informative, warranting the consideration and incorporation of alternative indicators in future research.

Land use. For land use, I explored three indicators, informal settlements (*inf_settl*), rural/urban (*rururb*) and macroareas (*zonas*). São Paulo is home to the largest slum population in South America, with approximately one-third of its residents living in slum-like conditions. Slums are often associated with low household income, inadequate awareness of heat and its correlation with health, poor living conditions and unstructured housing in below-standard conditions. Despite this, in this study, differences in risk between informal and formal settlements were marginal, and if anything, mortality risk was slightly higher for formal settlements. This contrasts with previous evidence, mainly from slums in Africa [335, 336] and India [337], and a study on impacts of heatwaves in Rio de Janeiro, Brazil[72]. In the studies from African and Indian slums, authors often describe dwellings as being built of metal sheets for both walls and roofing, and other lightweight material such as plastic and plywood. São Paulo's *favelas* predominantly have brick and cement foundations and walls, and use metal plates for roofing. Discrepancies in housing materials may account for the observed variations in outcomes. However, my stratified analyses based on the percentage of houses with non-conventional building materials (*build_matQ*) showed no substantial difference. Unfortunately, the variable had little variability with more than 80% of the areas with less than 3% of such dwellings. Another potential explanation involves differential exposure to temperatures between formal and informal settlements.

Silva and Ribero [338] investigated temperature differences between the *Paraisópolis favela* in São Paulo and a neighbouring area. The temperature in the *favela* was consistently 3°C higher during the day and 1°C higher during the night compared to the neighboring site. Short-term temperature variability (difference between maximum and minimum daily temperature) was also higher in the *favela*. These differences were more pronounced during extreme weather events. Prolonged exposure to higher temperatures and stronger extreme events, may induce long-term adaptation, potentially explaining the suggestive reduced risk of heat-related mortality here reported for informal settlements. However, when comparing the temperature distribution between formal and informal settlements as classified in my study, no statistically significant difference was observed, neither for cold nor for hot weather. Differences in within-day temperature variability (as opposed to daily mean temperature) were outside the scope of this study, yet warrant further investigation. Finally, differences in the classification of slum-like areas may explain the discrepancy of my results with the literature. My classification of formal and informal settlements was based on the São Paulo Urban Master Plan classification of areas. Thus, I defined informal settlements as those classified as ZEIS or vulnerable and socioeconomically deprived areas and in need of reform, which primarily includes slums, yet not exclusively. Even if only slum-like settings were included, in São Paulo these encompass three distinct types: *favelas*, which emerged as squatter settlements on public or private lands (n=1,600), "irregular" land subdivisions lacking legally recognized titles (n=1,100); and *cortiços*, which are overcrowded and poorly maintained tenement houses (n=1,900). Differences between these settlements may be relevant in the modification of the temperature-mortality association. Finally, the category of "formal settlements" includes any area not classified as informal setting resulting in an heterogeneous mix of population and areas thus, lacking specificity. Exploring the effects using a more accurate classification system that provides greater contrast would be insightful.

Between rural and urban areas, I found a higher cold-related mortality risk and lower heat-related mortality risk for rural areas. This is in line with some studies [16, 17, 339], yet evidence on this topic remains inconsistent, mainly due to methodological and classification differences, and most likely due to intrinsic heterogeneity between regions [340]. Demographic differences can be ruled out in my study as they are accounted for in the study design; thus, SES, access to services or behavioural patterns may partially explain these results, yet more research is needed.

Finally, regarding the macroareas, the risk estimates were only significant for the cold-effects. For those, the MCQUA, MEM, MQU and MUC were the areas with the largest risk estimates. These are all areas with some degree of urbanization and primarily located in the city center and its vicinity. Long-term exposure to higher temperatures due to the UHI, may lead to higher MMT and so, higher sensitivity to cold. In Gasparrini et al., 2017 [198] small-area study of England and Wales, despite they found higher MMT for large urban conglomerations, the cold-effects were lower than for rural areas.

Future research should investigate the characteristics of these macroareas in terms of socioeconomic, demographic and education and housing characteristics to disentangle underlying factors for this results.

Income. Evidence on the role of income inequalities is scarce, inconsistent and often using different indicators and categorization. The Gini index is a commonly used indicator. Here, I report a positive trend for heat-related mortality risk with increasing Gini index. This is in line with findings from a global multi-city study by Sera et al., 2019 [159], where the authors found high inequalities in income were associated to higher risk of heat-related mortality. However, this was not the case in Bakhtsiyarava and colleagues study[160], looking at 265 cities in Latin America where they found that cities with high Gini index had lower overall heat-related excess mortality compared to cities with low income inequality. This differences were not present when looking at extreme heat. Using a multi-city approach which included São Paulo, [341] found no statistically significant association for the Gross National Product per capita of the countries from which the cities were drawn.

Demographics. Areas with a higher proportion of elderly population showed slightly higher risk for heat and cold-related mortality. As discussed above, the elderly are more sensitive to non-optimal temperatures due to their higher prevalence of chronic diseases, social isolation, poor thermoregulatory system and behavioural differences with the younger population. Despite the strong evidence from individual level studies, findings are scarcer and more conflicting when it comes to area-level contextual indicators. For example, Sera et al., 2019 [159], using as indicator the area-level proportion of >65years old, did not find any significant difference neither for cold nor for heat. They suggest that it could be explained by the lack of range of variation in age distributions between areas and the higher proportion of elderly in rural areas. These could also be factors affecting my results. Regarding areas with different proportion of population belonging to different ethnic groups, I did not find any significant difference between groups, except for areas with more than 30% of population neither white nor black/mixed which showed a higher cold-related mortality risk. Nevertheless, the uncertainty was large and the contrast in the indicators used too small to draw strong conclusions. Racial segregation in Brazil has been diluting, yet its presence still remains [342]. The differences in risk observed here are likely to be a reflection of inequalities in education, income and living conditions [342]. This study, lacked the resources to explore the potential underlying factors, yet I strongly suggest that future studies investigate these inequalities in detail.

Employment. In general, the observed effects for heat were inconsistent, with only areas with low proportion of people with an income below once or twice the minimum wage showing a statistically elevated significant risk. For cold, the risk was significant for most groups yet no strong evidence of differences was found, except for a marginal lower cold-related mortality risk for areas with high unemployment rate and low legal employment. The evidence on area-level employment-related indicators

for temperature-mortality association studies is scarce [13, 17]. One of the very few studies available, that of Murage et al., 2020 [17] in London, showed no statistically significant difference in heat-related mortality, which would be in line with our findings.

Housing. Building characteristics play a crucial role in modifying temperature exposure and can serve as effect modifiers [151, 158]. For heat, Vandentorren et al., 2006 [151] found that during the August 2003 heatwave in France, elderly individuals residing in older constructions built before 1975 faced a higher risk of death (OR=1.83). Similar findings were reported in Zurich for individuals living in houses built pre-1970[67]. Heat loss through building fabric or heated air escaping can lead to challenges in adequately heating, especially in poorly insulated dwellings, which has been associated with excess winter mortality [158]. The overall findings of this study indicate a lack of significant evidence for an effect, and if anything, suggest a marginal decrease in the risk of cold-related effects in areas with a higher proportion of non-standard housing materials. Interestingly, the presence of overcrowding appears to be associated with a reduction in the risk of cold-related mortality, aligning with previous research [13]. However, it is important to note that even within the extreme quintiles, the prevalence of housing with these characteristics was minimal, thus limiting the ability to draw definitive conclusions.

SES Indices. The final contextual determinant group explored were SES composite indices. Among them, I included the widely used and globally available IDHM, to enable direct comparisons with other studies. Nevertheless, the global nature of the IDHM may not have fully capture the specific aspects of SES in São Paulo. To address this limitation, I also used the IPVS index, a vulnerability index specifically developed and validated for São Paulo. My findings seem to suggest that heat and cold-related effects were associated with fourth quintile of development when using the IDHM index and sub-component of education, but with lower development when looking and income and longevity components. Results from the IPVS index, showed contrasting results, with the heat- risk increasing with deprivation for both heat, yet the results were not statistically significant. This is in line previous evidence from São Paulo, where no strong evidence was found for the heat across SES groups [179, 343], even after stratifying by age groups [179]. However, it is important to acknowledge the difference in areal unit, time period and method used to generate the SES index by the different studies, which hinders direct comparison. In contrast, for cold, I did find some suggestive increase in the cold-related mortality risk with increased deprivation based on IPVS index. This is in line with findings from other regions, such as that from [16] in England and Wales; whereas Saucy et al., 2021 [67] only found an effect for heat-related mortality risk. The different SES indices used may explain the disparities in results. The SES indices are composite indicators aimed at providing a comprehensive and broad picture of the overall SES of a specific area. Yet, SES is a complex construct and vulnerability by SES is likely to vary by setting, outcome of interest, and exposure pathways. [179], included in addition

to income and education also variables relating to access to services and housing conditions in their index. Finally, my study used precise temperature exposure data which may explain differences. Finally, differences in the climate region and particularities of the area may also contribute to explain the diverging results.

5.4.4 Counter-intuitive results and comparability considerations

Some of the finding of effect modification by contextual determinants as examined in this study, contradicts previous evidence.

Firstly, it is possible that my findings are influenced by certain methodological limitations. The area-level indicators capture the characteristics of an area as a whole, but they may not accurately represent the experiences of all individuals residing in that area (known as ecological fallacy). Their was a lack of contrast in some variables with little variability across the study area, even when comparing extreme quintiles. For instance, the difference between an area with 5% of people at risk of poverty and an area with 10-20% may not result in a distinct enough differential impact to be detectable, thereby limiting power to detect a difference in effect size.

Alternative explanations proposed in the literature include the role of greenspaces and long-term adaptation [160]. Recent research by Moran et al., 2021[344] has shown that cities with low SES in Latin America often have a greater abundance of greenspaces. This may account for the lower rates of heat-related mortality found in cities with high poverty-related indicators, as reported by Bakhtsiyarava et al. (2023), [160]. This could explain the lower heat-related mortality risk observed with decreasing IDHM index, yet, it would not explain the higher cold-related mortality risk observed in the least deprived areas. The mechanisms underlying this effect modification could involve differences in exposure to greenspaces, e.g., cooling effect, as well as variations in the vulnerability of individuals based on their access to greenspaces, potentially associated with overall improvements in well-being. While the former mechanism may have influenced studies like Bakhtsiyarava et al., 2023 [160] where temperature allocation relied on coarse data, it is unlikely to significantly impact my results as my temperature dataset should adequately capture the cooling effect exerted by greenspaces. Exploring whether greenspace is associated with SES indicators used in this study and whether it has a role in modifying the observed differences is beyond the scope of this study, yet deserving attention for future work. Finally, Bakhtsiyarava et al., 2023 [160] showed that the poorest and most unequal cities in their sample were also the hottest, which had potentially led to long term-adaptation to higher temperatures in those cities. However, when comparing the temperature distribution across contextual determinant categories, I found no substantial difference.

SES encompasses a multifaceted structure that is shaped by various factors. Although the IPVS and IDHM indices make attempts to encapsulate its complexity, it is probable that they fail to account

for all dimensions of SES. Furthermore, these indicators are not specifically designed to assess the impacts of temperature on mortality, hence the pertinent factors associated with this outcome may not be adequately captured by these indices. Similarly, examining contextual indicators in isolation overlooks the combined influence of multiple conditions. This also includes the potential confounding or interaction with individual level characteristics. For example, different effects of SES have been reported by sex and age group [345]

It is also important to note that most of the evidence on the modification of the temperature-mortality association, specially by contextual determinants, are based on multi-city studies. These multi-city studies used meta-regression to investigate the association between the risk and the different contextual indicators, investigating all determinants at once in a multi-linear regression. This has important consequences for the interpretation of the results and when drawing comparisons with my findings as my analyses used a stratified approach. This means that each determinant is investigated separately and confounding is not accounted for. Also, determinants can be included in the meta-regression as continuous variables in the case of Sera et al., 2019 [159], or categories in the case of Bakhtsiyarava et al [160]. The former, may be masking some effects if non-linear associations are present; the latter, often results in study-specific categorizations of the factors which hinder direct comparison. Finally, multi-city studies inherently focus on comparing differences between cities rather than examining variations within each individual city. Thus, their findings may not translate to variations within a specific city, as it is the case of my study. This can be seen as a specific case of ecological fallacy. Finally, most studies examining individual determinants commonly include sex, age groups, education, and marital status, often using a comparable coding system, enabling meaningful comparisons. However, this consistency is not observed for less common or location-specific individual determinants, such as occupation, where evidence is scarce and classification methods vary across settings. Similarly, contextual determinants are explored using a diverse range of indicators, which often differ among studies or, when similar, are categorized differently. These differences may account for some of the heterogeneity observed between studies.

While most of these results are consistent with the literature and in line with previous assessments conducted in similar geographical and climatological areas [108, 179, 215], they should be interpreted with caution. The quantification of the effect is dependent on the chosen effect summary, based on a reference corresponding to the MMT. In this analysis, the measured excess was computed versus a constant MMT (23.29°C) corresponding to the MMT for the whole population over the study period. The statistical approaches varied across studies, with some early studies using less flexible approaches.

It is also possible that inequalities in the health burden of temperature could be better captured at earlier and less severe stages, for example through hospitalization rates. Similarly, it could be that daily mean temperature is not the best predictor of disparities in the mortality risk in this setting.

Xu et al., 2020 [346] conducted a comprehensive and detailed study on the impacts of short-term temperature variation in over 1800 cities in Brazil and found strong evidence supporting a higher risk of hospitalization for people living in cities with lower literacy rates, urbanisation rates, average household income, and GDP per capita were more vulnerable to hospitalisation, particularly significant in young and older people, and for hospitalisation due to infectious, respiratory, and endocrine disease. Studies investigating alternative temperature indicators such as short-term temperate variation and their impact on less severe health outcomes are needed.

Finally, the limited spatial variability observed in the modelled temperature estimates may account for the absence of discernible differences across contextual groups defined by spatial location. In Chapter 4, my analyses revealed a modest spatial variability of around 4°C in the modelled temperatures. This relatively constrained gradient might not exert a substantial influence on population health outcomes. Despite the model demonstrating high accuracy and performance metrics during both cross-validation and external validation, it is essential to acknowledge the possibility of temperature misclassification. While the robust validation results giving confidence in the model outputs, caution is warranted in interpreting temperature-related implications for contextual factors.

5.4.5 Strengths and novelty

Most importantly, my study was unique in the high spatial precision of health and temperature data. This is rare in temperature-mortality research where most studies use meteorological station [160, 179, 215], or temperature at coarser resolution, often above $1km^2$ [70, 159]. Here, I used daily mean temperature at $500m^2$ resolution assigned to each individual mortality record based on residential address, which reduces the risk of exposure misclassification. An important aspect to consider here, is that when using spatially resolved datasets, most of the effect modification that may occur through changes in the exposure pathways will be already captured in the model. This is the case for greenspaces, for example. In studies with coarse temperature resolution, greenspace is often suggested as a modifier for its cooling effect, which cannot be captured by the temperature data due to coarse spatial resolution (point location or coarse resolution). However, this is not the case in highly resolved datasets like mine, where exposure pathways are directly captured by the temperature. In this example, this means that pixels in and nearby greenspaces have lower temperature. Furthermore, other studies where strong effects were found, usually used monitoring stations to assign temperature. This may indicate that a large proportion of the effect may be mediated by differences in the exposure levels. Since these are accounted for in my model, the effect is reduced or deleted. Nevertheless, when comparing the temperature distribution between contextual factors, there was not a strong difference. Future work should investigate in detail potential differences in exposure levels.

A key point of distinction lies in the utilization of time-stratified case crossover design, which

confers numerous benefits over alternative methodologies. By adjusting for not only time-invariant, subject-specific confounding variables, such as sex, age, or educational status, intrinsic to the case crossover design, but also slow-variant confounders and temporal trends, including seasonality, long-term trends, and day-of-the-week effects, my study delivers a comprehensive and nuanced analysis of the data. Furthermore, the application of precise spatial and temporal individual exposures effectively minimizes the risk of exposure misclassification that may mar the reliability of other investigations in the field. Notably, the geo-referencing process, with a remarkable success rate of 95.25%, demonstrated the precision and rigor with which individual death records were linked to accurate residential address information.

Finally, the adoption of a flexible modelling approach to account for non-linear and delayed effects enhances the study's ability to capture the intricate relationship between temperature and mortality. The DLNMapproach in tandem with the case-crossover design, as recommended by Green et al., 2019 [12], proved instrumental in accounting for lagged effects and possible harvesting effect, a significant advancement in previous case crossover design studies, in line with recommendations from notable sources [179, 215]. In addition, the study benefits from an abundance of effect modifier data, particularly encompassing a wealth of socioeconomic and built environment contextual variables at a granular spatial scale. Notably, São Paulo boasts a rich repository of high-quality contextual indicators that are freely accessible through the local government website, further enhancing the comprehensiveness of the investigation.

5.4.6 Limitations

This study has certain limitations which deserve to be acknowledged. In Chapter 6 Section 6.5.1.4, I discuss some approaches to address some of these. Firstly, outdoor temperature as the exposure metric does not capture the totality of the expose profile of the individuals. Considering both indoor and outdoor temperature might better reflect personal exposure, as individuals spend significant time indoors. Murage et al., 2020 [17], in their simulations of indoor temperature, found that mortality was lowest in the quartile with the lowest indoor temperature and highest in the quartile with the highest indoor temperature. It is worth noting that indoor and outdoor temperature may not always be strongly correlated nor have a linear relationship [51]; in certain cases, factors such as building characteristics have a more substantial influence on indoor temperature variations than the location of the building within the city and the urban heat-island [347]. Moreover, people move within the city, so they are likely to be exposed to different micro-environments throughout the day, not just their residential address (e.g. work, commute...). Another limitation arises from the absence of spatially resolved air pollution data and RH, which limited the ability to account for these factors with an adequate spatial resolution. Although the use of the arithmetic mean of available air pollution stations

is common in the literature [179], it may smooth differences and result in exposure misclassification.

Finally, in order to ensure sufficient statistical power, the analyses were performed for each characteristic individually. This approach might oversimplify the relationships and obscure some genuine associations, particularly in the case of highly correlated covariates. The main motivation of the current study was two-fold: on the one hand, to demonstrate the usability of precise temperature and mortality data for epidemiological research, and on the other, to conduct a preliminary exploration of potential contextual and individual determinants influencing the temperature-mortality association. And thus, this approach was adequate. However, future research should consider stratifying the data using multiple determinants, both individual and contextual, at once (e.g., SES categories by age groups) to disentangle the findings and uncover more nuanced relationships. Similarly, the lack of information on physical and behavioral adaptation indicators, such as air conditioning use. Studies have shown that increased air conditioning use contributes to the attenuation of the temperature-mortality association in various regions [348]. Incorporating such adaptation indicators could provide valuable insights into the overall impact of temperature on mortality. Furthermore, the use of area-level summary indicators as contextual determinants, may not fully capture individual experiences and can lead to the "ecological fallacy" error.

5.5 Summary

In this chapter, I provided a comprehensive and extensive analyses of the effect modification of the daily mean temperature-mortality association by individual and contextual determinants in the municipally of São Paulo between 2015 and 2019 using spatially precise temperature and mortality data. To my knowledge, this is the first study to use spatially resolved temperature data to investigate the effects of cold and heat on mortality in Latin America. By employing spatially resolved temperature data, this study aims to minimize exposure misclassification, which represents a common limitation in similar studies. Overall, the effect of cold on mortality was stronger compared to the effect of heat, consistent with previous research conducted in this climatic region. The findings regarding individual determinants align with existing evidence, demonstrating an elevated risk of heat-related mortality among older individuals, women, and those with limited education (less than three years), although some of the estimates were not statistically significant. Regarding cold-related mortality, the effects were predominantly statistically significant, with a highest risk recorded among those of advanced age, lower educational attainment, mixed ethnic background, and married or widowed. In contrast, the findings for contextual determinants were inconclusive and spurious, with some unexpected results. Evidence on the effect modification of contextual determinants is scarce, heterogeneous and inconclusive, which makes comparison difficult. The lack of variation in some of the variables uti-

lized, the presence of ecological fallacy, confounding and/or interactions with other factors, as well as the spatial unit bias, may partially explain these varied and conflicting findings. Additional research is warranted to gain a better understanding of the influence of contextual determinants on the temperature-mortality association. Such knowledge would be valuable for informing both short-term and long-term adaptation measures.

Chapter 6

Discussion

6.1 Research in context

6.1.1 Evidence before this study

While there is a consensus on the impact of non-optimal temperatures — defined as any deviation from the temperature at which a population experiences the lowest mortality risk—on mortality, the precise shape and magnitude of the risk associated with heat and cold vary significantly across studies [3, 5–8, 12, 134]. Multi-city studies have played a vital role in determining whether these differences are genuine or merely artifacts resulting from variations in methods, temperature indicators, and thresholds used by different studies[3, 172]. By employing standardized methodologies and examining multiple cities across diverse climate regions, these studies have confirmed that the temperature-mortality association exhibits substantial variability across climatic regions [9, 10, 30]. In general, populations in tropical and subtropical climates demonstrate better adaptation to warm temperatures, whilst populations in colder climates exhibit larger heat effects [3, 9, 30, 299].

The observed differences in the health effects of heat and cold across various climate regions suggest that communities have adapted to their local climates over time. Furthermore, advancements in technology and changes in behavior have been identified as effective ways to reduce vulnerability to temperature. Understanding adaptation patterns in a population can help identifying the underlying mechanisms, and inform future preparedness strategies. Thus, a growing number of studies are investigating temporal trends in the temperature-mortality relationship. Fewer studies with less conclusive findings have reported evidence of adaptation to cold temperatures [86]. Disparities in adaptation capacity become apparent when examining different population groups, such as variations based on age and sex [84, 89–91]. However, evidence regarding the adaptation of other population groups, such as those defined by ethnicity or socioeconomic status, remains limited [86, 92, 93]. Finally, most of these studies, primarily conducted in developed countries, reported varying degrees of adaptation by

location [11, 40, 84–87], highlighting the importance of local assessments [13, 88].

When it comes to exploring potential vulnerable groups, the majority of research has primarily focused on individual determinants [13]. Age is the most extensively studied determinant, with evidence showing higher risks for heat and cold mortality among the elderly. Children and infants also exhibit higher risks, possibly due to physiological differences and behavioral patterns. Sex is another significant determinant, with women generally having higher mortality risks associated with heat, but findings for cold impacts remain inconclusive. Education, ethnicity, occupation, prior medical conditions, living alone or in isolation, and marital status have also been studied, but evidence of their role as effect modifiers varies, with limited and less conclusive results. This limited focus is mainly attributed to the lack of individual-level information routinely collected in death certifications.

The available evidence regarding the impact of urban characteristics and socioeconomic factors on the temperature-mortality relationship is limited, as highlighted out by Son et al.'s 2019 literature review [13]. One of the main challenges consists in the difficulty of accessing mortality data that is geographically referenced, which hinders the exploration of how contextual determinants may modify the association between temperature and mortality. Additionally, comparison between studies is hindered by the heterogeneous myriad of indicators used. Among the commonly investigated indicators are composite indices that capture socioeconomic inequalities[179], or individual indicators that are considered to act as proxies for such inequalities [17, 72, 151]. While there is a general trend suggesting that higher socioeconomic deprivation (and its proxies) is associated with an increased risk of temperature-related mortality, there are some studies that present conflicting results [179]. This lack of consistency has hindered the comprehensive understanding of temperature-related vulnerabilities, as well as the identification of vulnerable groups and areas that are particularly susceptible to heat and/or cold impacts.

Most studies investigating the temperature-mortality association rely on temperature data from a single meteorological station or the average temperature across all available stations[3]. This approach overlooks the spatial heterogeneity in temperature distribution, which can introduce bias in epidemiological findings [253]. Recent systematic reviews have highlighted the significant role of city micro-climates as effect modifiers of the temperature-mortality association [254]. Therefore, it is crucial to utilize temperature data that is spatially and temporally resolved to provide reliable and accurate evidence regarding the impacts of temperature on population health. However, accessing such data is often challenging or unavailable, particularly in LMIC.

Finally, although the majority of research on this topic is based in urban settings, with two large multi-city networks dedicated to this research (i.e., MCC [167] and SALURBAL [170]), the majority of studies tend to examine city-wide characteristics rather than exploring patterns within cities themselves, with a few exceptions [16, 173–175]. Cities are complex and heterogeneous systems

with various social inequalities, diverse built environments, and income and racial segregation, all of which are known to be determinants of health [156]. There is reason to believe that these factors impact the temperature-mortality association; however, evidence at the small-area level is lacking.

6.1.2 Added value of this thesis

This thesis offers a comprehensive investigation into the effects of non-optimal temperatures on mortality in São Paulo, Brazil. Chapter 3 examines the trends in the temperature-mortality association across population groups to understand the adaptation patterns present in the last two decades. In Chapter 4, I show the successful utilization of earth observation data and machine learning algorithms to generate high spatial resolution daily temperature estimates. Finally, in Chapter 5, I used this high resolution temperature exposure data to examine these associations for the first time in Latin America.

In Chapter 3, I investigated temporal trends in the cold and heat mortality risk over 15 years using a time varying DLNM model by cause of death, sex, age, and ethnic groups. The results indicate changes in the optimal temperature or MMT, suggesting shifts in the comfort temperature, as well as changes in the risk associated with extreme temperatures. Overall, the observed trends in MMT align with previous studies on adaptation in other locations. However, there is a slight increase in heat-related mortality relative risk over time. I also observed disparities in the presence and magnitude of adaptation across population groups are evident. Men, in general, experienced a reduction in heat-related mortality risk and an increase in cold-related risk over time, while the opposite effect was observed in women. Among the elderly population, the heat risk increased, mainly driven by an increase among women. Ethnic disparities revealed a slight decrease in risk for both heat and cold among the non-white population, primarily driven by those aged 65-79 years and males. Conversely, the white population experienced an increase in both heat and cold risk, consistent across age and sex groups. There was limited evidence of an association between MMT trends and annual mean temperature, as well as between mortality risk for heat and cold and extreme annual temperatures. These findings suggest that factors beyond climate play a role in the observed trends, including changes in risk awareness, socioeconomic factors, and access to services, among others. The results emphasize the need for regular monitoring of the temperature-mortality association among different population groups to inform appropriate public health interventions.

Despite advancements in climate science and epidemiological research, availability of spatially and temporally resolved temperature data remains limited, particularly in LMIC. In Chapter 4, I addressed this data gap by developing and implementing a RF algorithm to predict daily mean temperature at $500m^2$ resolution for the municipality of São Paulo. This approach combines information from ground monitoring stations and satellite-derived products. The RF model outperformed the traditional multi-

linear regression model and provided a more flexible framework to model non-linear temperature associations and handle complex interactions among input features. While highly resolved temperature models exists for many high-income countries , for example, UK [17], Switzerland [67, 69] and USA[306, 349], such models are virtually nonexistent for LMIC. My thesis demonstrates the feasibility of building statistical models to predict temperature with high accuracy in LMIC, serving as a blueprint for future research.

Spatially and temporally resolved models, such as the one developed in this thesis, hold great value for epidemiological studies for three key reasons. Firstly, they help to mitigate the measurement error introduced by the absence of precise data, which can lead to biased health effect estimates[350]. By accurately assessing exposure, this study validates existing evidence from studies that relied on single meteorological stations or coarse resolution temperature models, while minimizing the potential for misclassification bias. Secondly, spatially resolved temperature estimates are particularly valuable when examining potential pathways for effect modification. One such pathway is through alterations of the exposure levels, for example, the presence of greenspaces might lower local temperatures. This difference in temperatures due to the presence of greenspaces would be directly captured by the temperature model and the effect of greenspaces removed from the risk estimates. Residual effects could then be attributable to alternative pathways, such as adaptation resulting from prolonged exposures to higher or lower temperatures. Even if the immediate lower temperature experienced at the time of the event is captured by the model, people living near greenspaces may be used to living in lower temperatures and so, be more sensitive to heat, resulting in an increased risk for those areas. In combination with studies assessing exposure differential this can shed some light on the causal pathways behind the effect modification of certain contextual determinants. Finally, spatially and temporally resolved temperature models allow the exploration of risk profiles across population groups and small area characteristics, such as contextual determinants or spatial units (e.g., administrative units). I showcased this in Chapter 5, where I spatially linked mortality records to spatial characteristics based on their residential area at time of death. To date, this is the first study to comprehensively investigate a wide variety of individual and contextual determinants, which are thought to modify the mortality effect attributed to exposure to high and low temperatures in São Paulo using precise exposure estimates. This highlights the value of spatially resolved temperature data in LMIC settings, and shows new avenues for research to deepen our understanding and facilitate the identification of vulnerable groups and areas in urban settings.

6.2 Strengths and novelty

In my thesis, I have developed an open-access and downloadable spatiotemporally resolved temperature model, a stepping stone in mapping temperature exposure across the city which can be used to inform urban planning in São Paulo. The protocol and machine learning pipeline I designed for predicting ambient temperature in São Paulo daily at 500m^2 spatial resolution can serve as a guide for applying similar approaches in other cities. As discussed in Chapter 4, the model was designed and trained to capture the features of São Paulo and some modifications may be needed before models can be transferred to different contexts. With this approach, I showed the feasibility of building models which provide reliable temperature estimates. With the fast paced increase in satellite data availability and their improvement in resolution and quality, I expect future models to outperform the current one. Until then, my model provides what is currently the highest spatiotemporal resolved model of urban daily temperature for a Latin American city, and can support future environmental epidemiological research and local environmental decision-making processes in São Paulo.

Another significant strength of this thesis is the abundance and quality of data available for São Paulo, making it a compelling case study for investigating risk profiles across diverse groups and small areas. The mortality data encompasses a population of over 11.7 million individuals residing in the urban area, akin to the populations of some European countries. The mortality counts enable a robust exploration of population vulnerability without compromising statistical power. This benefited both Chapter 3 and 5, allowing me to explore temporal changes and differences in risk across population groups, respectively. Furthermore, it is noteworthy that the data was of very high quality, and the geo-referencing of the mortality records demonstrated a very high level of accuracy. For the temperature prediction (Chapter 4), I benefited enormously from the high-quality temperature data and the density of ground monitoring stations available in São Paulo, which sets it apart from other settings. Finally, in Chapter 5, the availability of high quality, diverse and abundant geo-referenced contextual data allowed me to explore several contextual characteristics which had never been explored before, including informal settlements which in most countries are not mapped nor included in the census. Since 2002, São Paulo includes informal settlements in their census, which offers a unique opportunity to investigate the health impacts on this specific population group characterized by distinct housing conditions, lifestyles, and occupational profiles.

Finally, the statistical models employed across my thesis are a notable strength, as they are considered state-of-the-art methods to assess the association between temperature and mortality. Both time-series and case crossover study designs have been shown to be the most adequate study designs to investigate short-term non-linear and delayed effects of temperature on mortality. Here, I coupled these with the flexible DLNM, which models the risk over the lag and exposure dimension.

6.3 Limitations

Despite the strengths and novelty of my thesis, there are some limitations worth discussing. Firstly, there is potential for residual confounding, inherent to all observational studies. In my analyses, I adjusted my models for the main confounders documented in the literature, namely air pollution and relative humidity. Unfortunately, confounders did not have the same data quality or spatiotemporal resolution as the temperature estimates, which may have led to some exposure misclassification. Moreover, emerging perspectives suggest that air pollution may act as an effect modifier, implying the need to explore temperature effects based on different air pollution categories. This aspect presents an interesting avenue for future investigations.

Secondly, there were specific limitations with regards to the temperature prediction model. Firstly, RF algorithm cannot predict temperature extremes exceeding the range of temperatures used in the training dataset. Although this limitation did not seem to significantly impact the current study's time period and area, it could potentially affect model performance when transferred to different time periods or settings. Additionally, model evaluation indicated reduced performance in rural settings due to a scarcity of training data. However, the extent of misclassification could not be accurately assessed due to inadequate validation stations. Finally, this model was tailored to capture temperature micro-conditions in São Paulo, and according to my model validation, it did so successfully. This high specificity can potentially limit its transferability to other settings. Developing a generalizable model was not an objective of this thesis, and thus, this is not a limitation but a consideration when and if attempting to transfer the model to other locations/times.

Finally, there were a few limitations with regards to the case-crossover analyses. Despite accurately allocating temperature exposure at the individual level using the developed temperature estimates and geo-coded mortality records, there remains a concern regarding residual exposure misclassification. This is because temperature was assigned based on residential addresses at the time of death, disregarding individuals' significant time spent at their workplaces or other locations. Additionally, indoor temperatures may differ notably from outdoor temperatures, leading to an unknown exposure misclassification, which varies depending on building types and conditions. Furthermore, although the study examined various contextual and individual characteristics, behavioral patterns influencing temperature exposure were not considered due to data limitations. Incorporating potential confounders in the model was restricted by data availability.

6.4 Public health and policy implications

Exposure to extreme heat in urban settings has increased by 200% between 1983 and 2016 [20]. The combination of increasing population exposure, susceptibility, and a warming climate presents a critical

environmental health challenge in the 21st century [171]. In Brazil, Armani et al. (2022) forecasted substantial increase in both minimum and maximum temperatures for the period 2020-2050, compared to the historical period of 1961-1990 of up to 5.0°C for RCP8.5 and 4.0°C for RCP 4.5. Additionally, the duration of heatwaves was projected to increase between 25 and 150 days for RCP 4.5 and up to 181 days for RCP 8.5. This was further supported by Feron et al. (2019) who projected a significant increase in heatwave duration, frequency, and number for the period 2046-55 compared to 1961-1990 across major South American cities, including São Paulo. Finally, according to a study by Kephart et al., 2021[171], the frequency of extremely hot days in South America's largest cities is predicted to increase five to tenfold by mid-century under a mid-level climate scenario (RCP4.5). These climatic changes pose significant public health threats, exacerbated by Latin America's aging and urbanized population. While there is a growing body of evidence on temperature-related health impacts in South America, there is a lack of research on intra-city variability of both exposure and vulnerable groups and neighbourhoods, and the capacity of communities to adapt, which is crucial for understanding localized risks and providing adequate responses [171, 179]. My thesis, has addressed both research gaps by (i) investigating changes over time in the risk (Chapter 3) and (ii) investigating population groups and neighbourhoods particularly at risk (Chapter 5) using a spatiotemporally resolved temperature model developed by myself to ensure precise temperature exposure allocation (Chapter 4)

Climate change strategies and plans are starting to recognize that the largest burden from temperature does not occur during extreme event periods; instead, is the moderate cold and heat which has the largest death toll in most populations [3]. To develop sustainable and climate-resilient cities, there should be a shift towards the development and implementation of strategies which include not only short-term emergency responses to extreme events but also long-term prevention plans that adapt the population to moderate and sustained temperatures. For this, it is essential to gain a better understanding of the contextual factors that contribute to vulnerability in urban settings. While some studies have examined individual risk factors, such as gender and age group, there is limited research on the impacts of contextual factors related to the natural and built environment, as well as the socioeconomic characteristics of urban areas. Expanding this knowledge, as I have done in my thesis, is useful to inform targeted interventions. Below, I discuss how this information can be used to inform both short-term and long-term interventions.

6.4.1 Short-term interventions

Surveillance systems are one of the most important tools to respond to climate hazards. In this context, HHWS have demonstrated their efficacy in reducing mortality associated with heat events. Ideally based on epidemiological evidence, these systems provide alerts to the population when temperatures are forecasted to reach health-threatening levels, thereby triggering the activation of national,

regional, or local plans to protect public health. As evidenced in Chapter 3, research highlights the temporal variability of risks, indicating the need for regular updates to these systems. Therefore, a systematic assessment of temperature-mortality risk is crucial to continually inform and enhance these HHWS. Finally, vulnerability to temperature is heterogeneous, conditioned by a myriad of individual and contextual determinants. These determinants are not static and may change over time, leading to variations in the vulnerability profile of a region. The identification of vulnerable groups and a comprehensive understanding of the mechanisms driving their susceptibility can provide valuable insights for targeted interventions aimed at at-risk populations. Moreover, this knowledge can aid in prioritizing specific vulnerable groups for inclusion in surveillance systems, ensuring a more effective and tailored response to climate-induced health risks.

6.4.2 Long-term interventions

The interplay between the built and natural environment significantly influences the health and well-being of individuals[351]. Notably, a recent systematic review conducted by Igle-Elegbede et al., 2022 [351], found that neighborhood design principles, such as walkability and access to greenspaces and amenities, are closely associated with community health and well-being. The effect of the built and natural environment on health effectively occur through alterations in the exposure pathways to certain hazards and/or the vulnerability of the population to their impacts. This is also the case for temperature and health and thus, long-term changes in the urban environment can help protect populations from the current and future temperature risks.

Accordingly, one of the primary strategies to mitigate the health risks posed by temperature is by reducing exposure to elevated temperatures. It is undeniable that the layout and design of cities significantly contribute to the magnitude of the UHI effect [14]. Therefore, retrofitting building designs and city planning to minimize heat accumulation and provide localized cooling stand as key long-term adaptations to rising temperatures in urban areas [14]. According to Raven et al., 2018 [352], four essential elements should be considered when planning and designing future cities: improving the efficiency of urban systems to minimize energy waste and greenhouse gas emissions, particularly from transport and housing; modifying the form and layout of cities to create green corridors and facilitate ventilation; using low heat capacity construction materials and reflective coatings; and increasing vegetation cover in the city while ensuring equitable access.

In addition to its impact on the UHI effect, urban planning significantly influences population distribution and, consequently, the distribution of vulnerable groups. Undesirable features in a neighborhood, such as inadequate road networks, poor water and sewage systems, and a lack of greenspaces, are often associated with poorer housing and living standards, attracting population groups with lower financial capital and socioeconomic status. In extreme cases, this results in distinctly demarcated ar-

eas of poverty concentration, with a high prevalence of vulnerable individuals. These inequitable neighborhood divisions can be traced back to historical events, as seen in examples like Redlining in the USA or the formation of *favelas* in Rio de Janeiro and São Paulo. Thus, altering the urban layout to avoid concentration of vulnerable groups in substandard neighbourhoods and reduce inequalities can also help manage temperature health risks.

Finally, mitigating the impacts of heat through urban planning not only protects the population from future temperatures, but also offers the opportunity to create healthier, more equitable, and thriving communities. Therefore, health co-benefits of urban planning interventions should be a central component of any proposed urban development [353]. This approach, encouraging systems-thinking and a comprehensive outlook with health at its core, is gradually being integrated into urban planning in recent years. For instance, the USA Environment Protection Agency promotes the concept of "Smart Growth" for their future urban developments, which encompasses various development and conservation strategies that safeguard health and the natural environment while enhancing community attractiveness, economic strength, and social diversity [354]. In São Paulo's Master Plan [213], though health is not explicitly addressed, the initiative promotes several activities with known co-benefits for health, such as improving housing conditions, reducing inequalities, and promoting active mobility.

In a commentary by MacLachlan and colleagues [355], earth observation was identified as a promising resource for providing information at high spatiotemporal scales to measure and monitor urban development for UHI mitigation within cities. Their study demonstrates how this data can be integrated with climate data and urban models to determine optimal tree placement and building layout to minimize the UHI effect. This is just one example of the potential of earth observation in urban planning and UHI mitigation. My temperature dataset at high spatiotemporal resolution over São Paulo offers a unique opportunity to explore inequalities in the exposure to extreme temperatures across the city and to identify areas of special risk that may need targeted urban planing interventions. Finally, the findings in Chapter 5 showcase how this data could be used to identify population groups and specific urban characteristics that are associated with increased risk and so, that require special attention due to their heightened vulnerability to heat and/or cold impacts.

6.5 Future work

In this section, I explore exciting avenues that have the potential to build upon the findings presented in this thesis. Additionally, I address the broader applicability of the results and the methodology employed. The discussion unfolds in two distinct subsections. The first subsection delves into research gaps and future possibilities of temperature modelling at a high spatio-temporal resolution. The second subsection centers on the opportunities and challenges in quantifying the temperature-mortality

burden in urban settings.

6.5.1 Temperature modelling

6.5.1.1 External validation

In my thesis, I validated the model using the stringent target-oriented CV proposed by Meyer et al., 2018 [286] as well as a hold-out dataset of five stations, making it a more comprehensive assessment than previous studies. Nevertheless, these validation approaches have some setbacks. Firstly, there is the issue with preferential sampling. That is, the meteorological stations used to validate the model are set up in prearranged locations to meet different needs other than model training and validation, e.g., monitoring of weather conditions at airports, or collecting temperature to calibrate air pollution measurements. Therefore, their location was not defined to include a representative sample of the different temperature micro-climates present in the study area. A clear example of this is the low density of meteorological stations in sub-urban and rural areas, particularly in the south. As a consequence, the data used both for training and validation is likely not capturing the complexity of spatial and temporal patterns present in the area. This is a common and intrinsic issue with data-driven modelling approaches and highlights the importance of prioritizing the expansion of monitoring networks of high quality, with a high sampling density and with a good and representative coverage of the climatic conditions present [166]. The ideal alternative, would be the implementation of a monitoring campaign purposely designed to capture such micro-climates, with good spatial and temporal coverage to validate the performance of the model in detecting the complexity of the system.

6.5.1.2 Improved input data

Earth observation, both remote sensing (e.g. satellite data) and sensor technologies (e.g. personal monitors), is rapidly developing, with data at increasingly higher spatial and temporal resolutions becoming available at a global scale. As this technology advances, we expect more data to become open access, new physical and chemical properties being monitored, and improvements in the data quality, temporal and spatial coverage and resolutions to occur. These prospective input features open the door for potential model improvements. Similarly, with the improvement of computational power we expect more climate models to become available at higher spatio-temporal resolutions. Although their resolution will most likely be still far off from the granularity here discussed they can be used as input data for the model.

Cloud cover is a common problem in earth observation optical data. Optical remote sensing makes use of visible, near infrared and short-wave infrared sensors on board of satellites, which capture the solar radiation being reflected off the Earth's surface at several (multi-spectral) or hundreds (hyper-spectral) spectral wavebands in the range of 300–2,400nm. Expectedly, these data are sensitive to

weather conditions and heavily affected by cloud cover. This hinders the quality of the data and its usability, particularly when the spatial and temporal resolution of the target variable is really high (on the range of meters and days, respectively). To address this, most studies including mine, impute the missing data during the data processing stage. Alternatively, some researchers have suggested the integration of multi-scale and multi-sensor data products in order to overcome this limitation [166].

Finally, Verdin et al., 2020 [21], noted that there has been a decline in monitoring stations. It is important to remember, that earth observation is, in no case, a substitute for monitoring stations. Monitoring stations are essential to train the models and ensure that we derive accurate air temperature and heat metrics from the satellite products. Crowd-sourced air temperature and citizen science have the opportunity to provide many data points for training [304, 356] and should be considered as an alternative to government and agency run monitoring stations, which can prove costly.

6.5.1.3 Modelling outside of the training space

Machine learning algorithms can learn complex associations between predictor variables and the target exposure which can then be used to predict the target exposure at unsampled points. The range of values a predictor can take at a given prediction area and period of interest are known as predictor space. When we train a model, we use the information of the predictors together with the ground truth. As the ground truth is point data, we extract the predictor values specific to those locations for each time point. In other words, we take a sample of each predictor's space. This sample will be the only thing the model will see and learn from. Consequently, if the values in the sample do not represent the full extent of the predictor space, there will be values which the model will never see, and so, will never learn from. Therefore, when we predict at those locations, the model is likely to over or underestimate the real value. This results in two issues: the specificity of the model to its area of applicability, and the challenge with modelling extreme values.

It is common practice to take the model performance indicators as representing the performance of the model across all the prediction area. The truth is, that the performance indicators and scores provided have been estimated using only the small sub-sample of the predictor space. And so, we can only be certain that they represent the model performance on the points for which ground truth data exists and not necessarily that at other locations. If the sampling is random, balanced, and captures the full extent of the predictor space without leaving gaps, the model performance will represent well the performance across all the prediction area, including at unsampled points. However, this is seldom the case, with most instances having clustered samples with clear imbalances and gaps in the predictor space distribution. This leads to two main issues. Firstly, areas with unseen predictor spaces are problematic since the algorithm has not been able to learn about their relationship with the target variable. Secondly, we cannot be sure that the model validation and accuracy metrics obtained from

cross-validation apply to them, as they could have completely different behaviours. From this follows that there is a need to identify areas with predictor values which fall outside the predictor space seen by the model to flag them as problematic. This would allow the delineation of an area for which we are certain that our model can reliably make predictions which adhere to the performance indicators provided. In the field of environmental exposure modelling, Meyer and Pebesma 2021 [357] coined the term Area of Applicability (AOA) to refer to this area. This follows previous work in the field of chemical modelling, species distribution modelling and soil mapping [358–360]. Further detail and algebraic notation for this method can be found in Meyer et al., 2018 [286]. The authors advocate for the AOA, or one such indicator, to be of mandatory reporting when reporting on model performance, allowing the users to make informed decision on the model suitability for their purposes. Adding the AOA map to my model evaluation and output would equip the users of my data withe the information required to make informed decisions on whether the model meets their needs.

By definition, RF algorithm cannot extrapolate outside the training range [316]. Several alternatives have been suggested to address this. Some authors have suggested modifications of the RF algorithm to include extremes. For example, Park et al., 2018 [361] proposed a new modelling approach for extreme RF quantile regression, which combined the flexibility of RF and circular transformation. Similarly, Gnecco et al., 2023 [362] coined the *extremal random forest*, which is a combination of RF quantile regression with theory of extrapolation. This is a growing field and more research is needed to validate these approaches. Alternatively, one can combine multiple models to leverage the strengths and limitations of each of them, the so called *ensemble models*. Any given model has strengths and limitations, which makes them appropriate for specific tasks and settings. In other words, each models can generate predictions which capture one of multiple probable states for a given real system. From this follows, that the combination of multiple unbiased probable states will lead to a more accurate and less biased prediction, supporting the idea of ensembles. In short, by combining multiple models we can leverage the strengths and limitations of each model, eventually reducing uncertainty, error and bias, and improving model accuracy and generalization. Mathematically, it has been shown that given N models, the error resulting from their combination will never exceed the expected error of the constituents models. This suggests, that the ensemble predictions will never have a larger error than those from its constituents. In the best case scenario, for a given set of N models, we could reduce the error by a factor of N, provided that the errors of each constituent model are uncorrelated. In practice, the errors are typically highly correlated, so the reduction in overall error is generally small, yet present. There is a wide myriad of methods used to combine these models, also known as *consensus algorithms*. The most simple is the arithmetic mean or median of all predictions generated by the models which been shown to improve the estimates in some occasions [363] yet not in others [309]. Slightly more complex approaches are those which include model filtering and a form of weighting

based on model performance. An example of this is the *weighted average algorithm*, where models are ranked by performance, only the top performing ones are included and their predictions weighted by the performance score to produce the ensemble average.

Future work should explore ensemble techniques capable of combining the flexibility of RF algorithms with the extrapolation capability of regression approaches such as generalized additive models. I believe this research direction has the potential to solve some of the issues with extreme value prediction and help advance this field. For more examples, refer to Marmion et al., 2009 [358] whom provides a comprehensive overview of some of the most common consensus algorithms and compares their accuracy for species distribution predictions.

6.5.1.4 Model Transferability

As discussed in the 2 section, there is a need in epidemiological studies for temperature data at a high spatio-temporal resolution. In this context, there is an interest in understanding the *model transferability* of my algorithm, which refers to the ability of an algorithm to be legitimately used to make predictions beyond the spatial and/or temporal extent of the data used to train the model. One example, would be to use the model to make predictions for a different city. There exist several challenges that must be considered when evaluating model transferability according to Yates et al., 2018 [364]. Below, I describe some of them in the context of my model.

Feature data access and quality. The performance of the RF model is highly sensitive to the input data as opposed to purely geostatistical methods such as kriging. Thus, access and quality of the input data are vital to ensure the model can be implemented in a new setting. Most of the input feature data used to train the model was open-access and available worldwide. Thus, access to data would not pose an issue. Satellite data has the great advantage of continuous coverage and longevity over field measurements. They also avoid the intrinsic problems associated with monitoring data such as temporal and spatial biases in data collection, insufficient and changing sample sizes and locations, poor quality or discontinuity in the data availability. Nevertheless, even with satellite and satellite-derived products, we can expect small variations in data quality and even discontinuities in data provision as satellites and measuring instruments age, break, are upgraded and/or get decommissioned.

Parsimonious models. Excessively complex models risk over-fitting the training data which eventually can lead to predictions being biased or being too specific to the reference system, and thus not transferable. Greater transferability is thus generally expected in parsimonious models with fewer predictors and less complex parametrizations. Here, I used FFS to select the subset of features that explained most of the variability, and avoid over-fitting. This reduced the initial pool of 24 features to 8 features. The favouring of temporally static and spatially dynamic features in traditional feature selection approaches is a common problem. These variables result in combinations of unique

station properties, which in practice behave as unique ID. As such, they are pivotal in describing the variability of training dataset, yet they will likely hamper transferability of the model both within the prediction boundaries and to new settings, as those unique combinations of properties are unlikely to concur again. To minimize this, I used a target-oriented cross validation approach when assessing the model performance of each feature combination. Refer to Section 4.3.2.1 for more detail.

Predictor space similarity. Model transferability relies substantially on the similarity between the predictor space use for training and that of the new target system. In other contexts, spatio-temporal distances between the training and the target setting have been used as surrogates. However, for temperature predictions other aspects may play a larger role in defining this similarity such as level of urbanicity or climatic classification group. The minimum level of similarity required to support transferable models, however, remains unknown. Several researchers have attempted to quantify this, with different approaches being developed [359, 360]. The AOA approach developed by Meyer et. al. 2021 [357] is one of such methods which aims to asses the transferability by quantifying the differences between the predictor space used to train the model and that that of a target area. In section 6.5.1.3, I described how this method can be used to establish the area of applicability of the model. However, this can easily be extended to other target areas where the model is to be transferred.

Temperature-feature associations must mostly hold. The success of the model transfer inherently relies on the relationships describes by the model holding in the new environment. In other words, the model is trained to capture at the best of its capabilities the underlying complex associations that exist between input features and temperature. Thus, the model transferrability will be hindered if any of these associations change in shape or magnitude, if they do not exists in the new environment, if new associations, and/or if interactions between existing or new features are present.

Validation and re-training. Many factors influence the final transferrability of a model, several of which are hard if not impossible to assessed due to lack of information or uncertainty. Therefore, it becomes vital to validate the transferred model with ground truth data from the target location or time of the transfer. Of course, this is not always possible as meteorological stations may not be available in those settings. Moreover, bias in the sampling or quality of these validation points may hinder our ability to assess the model objectively, and should be considered when making a final decision on the model transferability. Finally, if there is a dense network of data points in the target location/time, it is possible to use a sample of these points to re-train the model. This process is less computationally expensive that the initial model training and would allow the model to capture new or slightly different associations that may be specific of the new setting. My model, was calibrated for a specific area, São Paulo, for a specific period of time, between 2015 and 2019. For this, it is unclear whether this model could be transferred to other settings. I believe there is value in assessing the transferability of the model to other settings as it could yield to temperature data being accessible in areas where no such

data exists. Future work should focus on investigating the transferrability of the model. For a given candidate location and based on the challenges described above, the following assessments should be considered: (i) ensure that all the feature data is available, of quality, covering the spatio-temporal extent of interest and at the appropriate resolution; (ii) define the predictor space similarity between São Paulo and the new location, potentially using an adaptation of the AOA approach from Meyer et. al. 2021 [357]; (iii) investigate whether threats to the temperature-feature associations may exists (e.g., different climate region, urban vs rural), and finally (iv) test the performance of the transferred model with a sufficient network of local ground monitoring stations that is representative of the area.

6.5.1.5 Operational EO for heat monitoring

Heat exposure assessments using meteorological station based approaches might fail to accurately capture geographic variation in temperature-related exposure risks, resulting in biased risk estimates. This can lead to exposed population not being correctly identified [166, 365]. Often high risk hotspots, such as remote areas with structurally-marginalized communities, have poor or no representation of temperature monitors, and using measurement approaches these areas are often ignored or misclassified. This potentially leads to HHWSs systematically neglecting these areas of higher vulnerability. This lack of representation also translates into underestimation of the benefits of climate action on mitigation and adaptation, as the hotspots where they are most effective are often those with poorer representation of temperature data and might be missed. Several aspects need to be addressed for an operational heat monitoring and or forecasting system to be successful.

Firstly, it should be dynamic when it comes to accommodating updates in the input data. That is, it should be able to absorb and integrate relatively fast any new relevant earth observation products available and updates or improvements of existing ones. Earth observation is a rapidly growing field. Its broad applications and usability has increased the demand and consequently, the capital investment. In the last decades, we have seen an exponential growth in the private sector, primarily with small startups bringing a field previously dominated by international and national space agencies, to a competitive commercial ground. This market competitiveness has and is likely to continue speed up the progress and advances on remote sensing technology. The approach most of these startups take is that of small satellite constellations with a modular built which allows for a versatile and multi-purpose use. One such startup is Ororatech in Munich, Germany funded by the European Space Agency (ESA) [366]. They provide IR thermal imagery at a resolution of 80m with a target temporal resolution of one image every 30 minutes of every place on Earth. Their products have already been used for wildfire detection and monitoring alert systems (e.g., British Columbia). During my stay at ESA, I discussed potential usability of this LST data for integration into a heat warning system. Although, these start-ups are providing promising and exciting new data, the limitation is often its

accessibility due to its commercial nature. Thus, international agencies should continue to invest in the development of such technologies and provide free access to the date for the public good.

Secondly, the development and establishment of efficient and rapid workflows to integrate satellite imagery into operational heat forecasting and warning systems that can act in real-time. This has intrinsic computational and technical challenges that will require of ingenious technical fixes, multidisciplinary collaborations and financial investment. Thirdly, such systems should combine the technical expertise of the different players involved, including remote sensing experts, climatologists, urban planers and public health experts. There are some examples of successful collaborations and programs that aim at bridging the gap between the earth observation-community and other fields, such as the Global Health Information Network, WMO or the Geo Health Community Practice. Another good example of collaborative and ambitious work is the "Destination Earth (DestinE): a European Commission flagship initiative for a sustainable future" led by the ESA[367]. This virtual model of the Earth should be able to recreate the complex functioning of the Earth systems by integrating multi-source and multi-scale data including earth observation products and mathematical models.

6.5.1.6 Interdisciplinary collaboration

The collaboration between various fields, such as air pollution modeling, economics or meteorological sciences, holds immense potential for enhancing the depth and breadth of research in the temperature-health domain. In the context of the mentioned thesis, there exists a substantial body of literature in air pollution studies that remains untapped. For instance, the observation that temporal predictions outperform spatial predictions in spatio-temporal models aligns with findings in pollution analysis described elsewhere [368, 369]. The impacts of measurement error on impact estimates, is another example of a topic more explored in air pollution modelling [370]. Climate-health scientists stand to gain valuable insights from this and other research, underscoring the relevance and potential benefits of cross-disciplinary knowledge exchange. This interdisciplinary approach not only enriches the depth of research but also promotes innovation and opens avenues for novel solutions to pressing challenges in the realm of air quality and pollution studies.

6.5.2 Quantification of the temperature-attributable health burden

6.5.2.1 Untangling the mechanism behind adaptation

Adaptation is a consequence of several mechanisms: physiological, behavioural, infrastructural, and technological. In order to understand the importance of each of these mechanism in the observed adaptation we need multi-disciplinary approaches. For example, Newsome et al., 2011 [240] followed 29 households in Detroit, USA, during the summer months. The residents were asked to keep an hourly log of eight heat-adaptive behaviors: opening windows/doors, turning fans or the air condi-

tioner on, changing clothes, taking a shower, going to the basement, the porch/yard, or leaving the house. Information on indoor and outdoor temperature, housing type and percent of nearby surface imperviousness was collected. They found a high correlation between heat-adaptive behaviors and indoor temperature, with a preference for mechanical adaptations (i.e., fans and air conditioning). These behaviours were more frequent among residents of high rising buildings with high percentage of impervious surfaces nearby. This is a good example of a comprehensive assessment of behavioural (e.g., showering), infrastructural (e.g., housing type and imperviousness) and technological adaptive mechanism (e.g., air conditioning and fans) and exposure levels (indoor vs outdoor temperature). The main limitation of studies such this, is their little transferability to other settings. Nevertheless, having a larger sample of studies with a variety of behaviours and different settings, potentially including measures of other confounders and effect modifiers such as air pollution and relative humidity or individual level characteristics of the respondents, would shed light in our understanding.

6.5.2.2 Need for collaborative, systems-thinking and target-driven research

Building temperature-proof and resilient cities that both protect their population from the temperature-health risks and mitigate their contribution to climate change is essential. In the context of urban policies, health promotion aligns well with climate change mitigation and adaptation strategies; however, this valuable synergy is often overlooked. Cities tend to adopt fragmented and isolated approaches to policy development. As a result, integrated solutions that address major public health risk factors, preserve the environment, and tackle climate change remain underutilized.

Urban planning is increasingly acknowledged as a crucial determinant in addressing various health, ecological, and climate-related challenges. Unfortunately, planners are not always equipped with the necessary knowledge or given the mandate to pursue health promotion, nature conservation, or climate mitigation and adaptation goals. This missed opportunity hampers the effective integration of health, environmental, and climate considerations in urban development. Hence, a more comprehensive research approach with a systems thinking perspective is required to capture relevant factors and impacts associated with specific challenges. Moreover, research with clear policy or urban planning outcomes would be beneficial, emphasizing collaboration and co-creation with end users, such as policymakers and urban planners. I discussed this in detail in a commentary [353], together with colleagues from the International Society for Environmental Epidemiology (ISEE).

6.5.2.3 Exploring spatiotemporal patterns

Small area estimates can be a useful indicator for local health behavior and disease surveillance purposes, and they have also been used in secondary analyses to identify potential community-level factors that may be associated with health behaviors or disease prevalence [371]. Thus, the availability

of granular data enhances the efficiency of the decision-making process because they can be used as the basis in formulating policies, identifying appropriate population groups for policy targets, and monitoring the progress of an already implemented program [371].

Temperature-related mortality risks display a considerable geographical heterogeneity. Thus, spatial assessments of temperature-related mortality can proof incredibly useful for informing, targeting, and implementing area-level public health policies and in future epidemiological analyses, as well as for projecting risks and impacts at small-area level under climate-change scenarios. To date, only few studies have investigated the temperature-related mortality risk at a small geographic scale in urban areas [17–19, 175] and only a few studies have explored larger regions [16, 173, 174, 177].

Advancements in three key domains are expected to facilitate the ease and prevalence of small area studies: enhanced computational capabilities; increased availability of high spatially-resolved health, exposure, confounder, and effect modifier data, and improvements in statistical methodologies. In line with this premise, the present thesis aimed to contribute to the expansion of access and availability to high spatially resolved temperature data in the São Paulo region, an area where such data has been lacking until now. Chapter 5 of this thesis demonstrated the utility of this newly available data by conducting intra-city vulnerability temperature-mortality assessments. Nevertheless, the ultimate ambition of this research lies in establishing a foundation for future small-area investigations in the region, thereby fostering a more robust and comprehensive understanding of localized impacts and vulnerabilities. By addressing the data scarcity issue and paving the way for future studies, this research strives to facilitate evidence-based decision-making and policy formulation to address health and environmental challenges more effectively in the São Paulo area.

6.5.2.4 Detailed investigation of the temperature-related health burden

Cold and heat have been associated with multiple diseases including: mental and nervous disorders, cardiovascular, hypertensive, respiratory, endocrine, genitourinary, digestive and infectious diseases, and external causes (i.e., accident or suicide). Most studies have examined the effects on CVD and respiratory diseases related mortality, and subcategories of those [67, 70, 138, 141, 162, 168, 171, 179, 372] with only a few investigating other diseases, including large disease groups [161], suicide [82, 334, 373] or cancer [374] or [161, 334]. In 2021, the Global Burden of Disease study published for the first time global cause-specific temperature attributable burden estimates for daily temperature [42]. However, only a few studies have explored how estimates differ by individual factors [67, 72, 80] and even less of contextual factors [67, 72]. It is likely that some effect modifiers act differently by causes of death, and thus, this warrens further investigation.

In addition, the physiological effects of ambient temperatures exhibit a wide range of consequences, extending beyond mortality to include intermediate health events. Therefore, further inves-

tigations into these intermediate health events are essential to fully comprehend the overall health burden. Understanding these early signs of temperature-related impacts holds significance not only from a clinical standpoint but also in terms of resource allocation and prioritization. With climate change progressing, extreme heat and cold events are expected to increase, leading to increased demands on the healthcare system. A comprehensive systematic review by Mason et al., 2022 [375] highlighted that such events would result in heightened requirements for ambulance services, emergency departments, hospital admissions and beds, as well as increased power and water consumption. This necessitates local services to assess their resource needs based on the demographic composition and prevalent medical conditions in their communities. While the focus of the review was on Australia, the findings hold applicability to various settings. Moreover, the economic impact is also significant, given the potential loss of working days, particularly in occupations involving outdoor exposure and physical exertion. Unfortunately, such studies are scarce, primarily focusing in hospital admissions, and mostly located in the northern hemisphere [376].

Countries with comprehensive and robust hospital and registry data offer valuable opportunities for investigating non-fatal outcomes and exploring potential causal pathways. For instance, a recent study conducted in Anqing, China [377] demonstrated a possible association between temperature and increased hospital admission risk among patients with rheumatoid arthritis. Cohort studies, on the other hand, provide a unique prospect to examine mild temperature-related outcomes and track their progression within a well-informed group of individuals, where underlying conditions and vulnerability factors are extensively documented. For example, Sun et al. 2017 [378] conducted a study in China on a prospective cohort of 66,820 elderly individuals (65 years) with 10–13 years of follow-up. Their research revealed that a 1°C increase in wintertime temperature variability was associated with a 1.20 (95% CI 1.08 to 1.32), 1.15 (1.01 to 1.31), and 1.41 (1.15 to 1.71) fold increase in the risk of emergency hospital admission for total respiratory diseases, pneumonia, and COPD, respectively.

6.5.2.5 In-hospital mortality

In-hospital heat-related deaths have been reported to increase during heat events [138, 153, 379], being considered by some as a nosocomial disease[153]. These has important implications. Current public health initiatives, including HHAP, focus on providing resources and support to susceptible individuals in their home and work environments, less so towards people who are already under medical care in hospitals. Some hospitals have plans in place to minimize those events yet it is not often part of the HHAP. Even though the mortality records used in this study had information on the location of death, including in hospitals, it did not include information on whether this was part of a longer stay or if the individual was hospitalized due to the eventually fatal heat-related event.

6.5.2.6 Artificial neural networks

Benefits of neural networks include: no prior assumption about the type of relationship, nor on the presence of interactions between variables; they can handle multi-collinearity without requiring any prior treatment of the data; they can handle large datasets. Leiriao et al., 2023 [380] investigated the usability of Artificial Neural Networks to predict CVD mortality among the elderly between 2007 and 2019 in São Paulo. For this study, the authors used a neural network with six input layers including PM_{10}, SO_2, NO_2, temperature, RH and wind speed, one hidden layer with 25 neurons, and one output layer. Splitting the models by season resulted in improved model performance and reduced model error. Variable importance varied substantially by season. For example, daily mean temperature was the second most important variable in explaining the CVD mortality in elderly for the rainy season whereas it seemed to play less of a role during the dry season, being classed fifth after all pollutants and RH. Thus, artificial neural networks can be an interesting tool not only to predict (and potentially forecast) the health burden of temperature, but also to investigate complex interactions and factors influencing it.

6.6 Conclusions

São Paulo, like many other cities, is expected to confront escalating climate change impacts, with temperature variations posing a significant threat. This thesis offers a comprehensive investigation into the temperature-mortality risk, examining temporal changes and with spatial precision.

The findings reveal that the temperature-related mortality risk in São Paulo has undergone substantial changes over recent decades, exhibiting heterogeneity across population groups and temperature ranges. This diversity suggests that various mechanisms underlie the observed trends, with differing levels of influence on the population. The absence of a discernible association between adaptation and temperature changes during the study period implies that external factors, such as behavioral, infrastructural, or socioeconomic changes, might predominantly account for the observed risk trends. However, further research is warranted to elucidate the precise mechanisms driving these trends. Despite observing reductions in heat-related risk for certain groups, a residual risk persists, emphasizing the necessity for continued efforts to mitigate peak mortality events associated with temperature. Additionally, the study highlights the significant impact of cold temperatures in São Paulo compared to other settings, a factor that should not be overlooked in the development of interventions.

A prevailing issue in most intra-city temperature-mortality studies is exposure misclassification due to the lack of spatiotemporally resolved temperature data. Addressing this crucial limitation, I devised and constructed a machine learning pipeline capable of predicting daily temperatures at a resolution of $500m^2$. The results demonstrate that the RF framework achieves highly accurate

and unbiased predictions, rivaling or even surpassing traditional regression methods. This novel dataset represents the first-ever open-access resource to provide daily mean temperature estimates at such a high spatial resolution for São Paulo (or any South American city), presenting unprecedented possibilities for epidemiological research on the health burden of temperature in this region.

I utilized spatially resolved temperature data to examine the effects of cold and heat on mortality, aiming to minimize exposure misclassification. This represents a pioneering effort in Latin America. The results reveal that the elderly, women, and individuals with lower education levels generally experience higher temperature-related mortality, particularly concerning cold temperatures. However, findings related to contextual characteristics, including economic and built environment factors, presented less distinct and more inconsistent outcomes. This study primarily aimed to showcase the potential of the temperature dataset in epidemiological research and serve as a stepping stone for future investigations. Additional research is warranted to gain a better understanding of how contextual determinants influence the temperature-mortality association, which would be instrumental in shaping both short-term and long-term adaptation measures.

In conclusion, my thesis underlines that the risks associated with heat and cold-related mortality remain a pressing public health concern, particularly for vulnerable groups, and their effects exhibit dynamic variations over time and space. Consequently, policy interventions should incorporate these findings into planning and implementation processes, including the development of local adaptation strategies and the establishment of thresholds and target groups for HHWS and HHAPs. Furthermore, with rising temperatures, equitable city-level actions and targeted interventions in urban planning should be considered to mitigate the current and future heat effects effectively.

Bibliography

[1] IPCC. *Climate Change 2014: Mitigation of Climate Change: Working Group III Contribution to the IPCC Fifth Assessment Report.* 2014.

[2] Ken Parsons. *Human thermal environments: The effects of hot, moderate, andcold environments on human health,comfort, and performance, third edition.* CRC Press, 1 2014.

[3] Antonio Gasparrini, Yue Liang Leon Yuming Yue-Liang Leon Guo, Masahiro Hashizume, Eric Lavigne, Antonella Zanobetti, Joel Schwartz, Aurelio Tobias, Shilu Tong, Joacim Rocklöv, Bertil Forsberg, Michela Leone, Manuela De Sario, Michelle L Bell, Yue Liang Leon Yuming Yue-Liang Leon Guo, Chang-fu Fu Wu, Haidong Kan, Seung-Muk Muk Yi, Micheline De Sousa Zanotti Stagliorio Coelho, Paulo Hilario Nascimento Saldiva, Yasushi Honda, Ho Kim, and Ben Armstrong. Mortality risk attributable to high and low ambient temperature: a multicountry observational study. *The Lancet*, 386(9991):369–375, 7 2015.

[4] Xiaofang Ye, Rodney Wolff, Weiwei Yu, Pavla Vaneckova, Xiaochuan Pan, and Shilu Tong. Ambient temperature and morbidity: A review of epidemiological evidence. *Environmental Health Perspectives*, 120(1):19–28, 2012.

[5] Rupa Basu and Jonathan M. Samet. Relation between elevated ambient temperature and mortality: A review of the epidemiologic evidence. *Epidemiologic Reviews*, 24(2):190–202, 2002.

[6] R Sari Kovats and Shakoor Hajat. Heat stress and public health: a critical revie. *Annual review of public health*, 29:41–55, 2008.

[7] Asya Dimitrova, Vijendra Ingole, Xavier Basagaña, Otavio Ranzani, Carles Milà, Joan Ballester, and Cathryn Tonne. Association between ambient temperature and heat waves with mortality in South Asia: Systematic review and meta-analysis. *Environment International*, 146, 2021.

[8] Talila Perry, Uri Obolski, and Chava Peretz. The Association Between High Ambient Temperature and Mortality in the Mediterranean Basin: a Systematic Review and Meta-analysis. *Current Environmental Health Reports*, 10(1):61–71, 2023.

[9] Aurelio Tobías, Ben Armstrong, and Antonio Gasparrini. Investigating uncertainty in the minimum mortality temperature. *Epidemiology*, 28(1):72–76, 2017.

[10] Qian Yin, Jinfeng Wang, Zhoupeng Ren, Jie Li, and Yuming Guo. Mapping the increased minimum mortality temperatures in the context of global climate change. *Nature Communications*, 10(1):1–8, 2019.

[11] Francesca K. de' Donato, Michela Leone, Matteo Scortichini, Manuela De Sario, Klea Katsouyanni, Timo Lanki, Xavier Basagaña, Ferran Ballester, Christofer Åström, Anna Paldy, Mathilde Pascal, Antonio Gasparrini, Bettina Menne, and Paola Michelozzi. Changes in the effect of heat on mortality in the last 20 years in nine European cities. Results from the PHASE project. *International Journal of Environmental Research and Public Health*, 12(12):15567–15583, 2015.

[12] Hunter Green, Jennifer Bailey, Lara Schwarz, Jennifer Vanos, Kristie Ebi, and Tarik Benmarhnia. Impact of heat on mortality and morbidity in low and middle income countries: A review of the epidemiological evidence and considerations for future research. *Environmental research*, 171:80–91, 4 2019.

[13] Ji Young Son, Jia Coco Liu, and Michelle L. Bell. Temperature-related mortality: A systematic review and investigation of effect modifiers. *Environmental Research Letters*, 14(7), 2019.

[14] Shilu Tong, Jason Prior, Glenn McGregor, Xiaoming Shi, and Patrick Kinney. Urban heat: An increasing threat to global health. *The BMJ*, 375:1–5, 2021.

[15] A. John Arnfield. Two decades of urban climate research: A review of turbulence, exchanges of energy and water, and the urban heat island. *International Journal of Climatology*, 23(1):1–26, 2003.

[16] Antonio Gasparrini, Pierre Masselot, Matteo Scortichini, Rochelle Schneider, Malcolm N. Mistry, Francesco Sera, Helen L. Macintyre, Revati Phalkey, and Ana Maria Vicedo-Cabrera. Small-area assessment of temperature-related mortality risks in England and Wales: a case time series analysis. *The Lancet Planetary Health*, 6(7):e557–e564, 2022.

[17] Peninah Murage, Sari Kovats, Christophe Sarran, Jonathon Taylor, Rachel McInnes, and Shakoor Hajat. What individual and neighbourhood-level factors increase the risk of heat-related mortality? A case-crossover study of over 185,000 deaths in London using high-resolution climate datasets. *Environment International*, 134(July 2019):105292, 2020.

[18] David M Hondula, Robert E Davis, Joacim Rocklov, and Michael V Saha. A time series approach

for evaluating intra-city heat-related mortality. *Journal of epidemiology and community health*, 67(8):707–712, 8 2013.

[19] Yihan Xu, Payam Dadvand, Jose Barrera-Gómez, Claudio Sartini, Marc Marí-Dell'Olmo, Carme Borrell, Mercè Medina-Ramón, Jordi Sunyer, and Xavier Basagaña. Differences on the effect of heat waves on mortality by sociodemographic and urban landscape characteristics. *Journal of Epidemiology and Community Health*, 67(6):519–525, 2013.

[20] Cascade Tuholske, Kelly Caylor, Chris Funk, Andrew Verdin, Stuart Sweeney, Kathryn Grace, Pete Peterson, and Tom Evans. Global urban population exposure to extreme heat. *Proceedings of the National Academy of Sciences of the United States of America*, 118(41):1–9, 2021.

[21] Andrew Verdin, Chris Funk, Pete Peterson, Martin Landsfeld, Cascade Tuholske, and Kathryn Grace. Development and validation of the CHIRTS-daily quasi-global high-resolution daily temperature data set. *Scientific Data*, 7(1):1–14, 9 2020.

[22] United Nations. Department of Economic and Social Affairs. Population Division. World Urbanization Prospects: The 2018 Revision (ST/ESA/SER.A/420). Technical report, New York: United Nations, 2019.

[23] IBGE. IBGE — Portal do IBGE — IBGE, 2020.

[24] Gabriela Narcizo de Lima and Víctor Orlando Magaña Rueda. The urban growth of the metropolitan area of Sao Paulo and its impact on the climate. *Weather and Climate Extremes*, 21(March):17–26, 2018.

[25] AinaRB (Aina Roca-Barcelo) · GitHub.

[26] Aina Roca-Barcelo. Climate and health in urban areas: the case study of Sao Paulo Brazil - Climate health burden.

[27] NCEI; NOAA; NASA. Assessing the Global Climate in 2022 — News — National Centers for Environmental Information (NCEI).

[28] NASA Earth Observatory. World of Change: Global Temperatures.

[29] V Masson-Delmotte, P Zhai, H-O Pörtner, D Roberts, J Skea, P.R. Shukla, A Pirani, W Moufouma-Okia, C Péan, R Pidcock, R Connors, J.B.R. Matthews, Y Chen, X Zhou, MI Gomis, E Lonnoy, T Maycock, M Tignor, and T(eds.) Waterfield. Global Warming of 1.5°C.An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change. *In Pres*, (October 2018), 2018.

[30] Qi Zhao, Yuming Guo, Tingting Ye, Antonio Gasparrini, Shilu Tong, Ala Overcenco, Aleš Urban, Alexandra Schneider, Alireza Entezari, Ana Maria Vicedo-Cabrera, Antonella Zanobetti, Antonis Analitis, Ariana Zeka, Aurelio Tobias, Baltazar Nunes, Barrak Alahmad, Ben Armstrong, Bertil Forsberg, Shih Chun Pan, Carmen Íñiguez, Caroline Ameling, César De la Cruz Valencia, Christofer Åström, Danny Houthuijs, Do Van Dung, Dominic Royé, Ene Indermitte, Eric Lavigne, Fatemeh Mayvaneh, Fiorella Acquaotta, Francesca De'Donato, Francesco Di Ruscio, Francesco Sera, Gabriel Carrasco-Escobar, Haidong Kan, Hans Orru, Ho Kim, Iulian Horia Holobaca, Jan Kyselý, Joana Madureira, Joel Schwartz, Jouni J.K. Jaakkola, Klea Katsouyanni, Magali Hurtado Diaz, Martina S. Ragettli, Masahiro Hashizume, Mathilde Pascal, Micheline de Sousa Zanotti Stagliorio Coélho, Nicolás Valdés Ortega, Niilo Ryti, Noah Scovronick, Paola Michelozzi, Patricia Matus Correa, Patrick Goodman, Paulo Hilario Nascimento Saldiva, Rosana Abrutzky, Samuel Osorio, Shilpa Rao, Simona Fratianni, Tran Ngoc Dang, Valentina Colistro, Veronika Huber, Whanhee Lee, Xerxes Seposo, Yasushi Honda, Yue Leon Guo, Michelle L. Bell, and Shanshan Li. Global, regional, and national burden of mortality associated with non-optimal ambient temperatures from 2000 to 2019: a three-stage modelling study. *The Lancet Planetary Health*, 5(7):e415–e425, 2021.

[31] The United Nations Office for Disaster Risk Reduction (UNISDR). Sendai Framework for Disaster Risk Reduction 2015 - 2030. Technical report, 2015.

[32] World Health Organization(WHO) & World Meteorological Organization (WMO). *Heatwaves and Health: Guidance on Warning-System Development.* Number 1142. 2015.

[33] Omar Dario Cardona, Maarten K. Van Aalst, Jörn Birkmann, Maureen Fordham, Glenn Mc Gregor, Perez Rosa, Roger S. Pulwarty, E. Lisa F. Schipper, Bach Tan Sinh, Henri Décamps, Mark Keim, Ian Davis, Kristie L. Ebi, Allan Lavell, Reinhard Mechler, Virginia Murray, Mark Pelling, Jürgen Pohl, Anthony Oliver Smith, and Frank Thomalla. Determinants of risk: Exposure and vulnerability. *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation: Special Report of the Intergovernmental Panel on Climate Change*, 9781107025:65–108, 2012.

[34] Renee A. Catullo, John Llewelyn, Ben L. Phillips, and Craig C. Moritz. The Potential for Rapid Evolution under Anthropogenic Climate Change. *Current Biology*, 29(19):R996–R1007, 2019.

[35] Beker Braian M, Cervellera Camila, Vito Antonella De, and Musso Carlos G. Human Physiology in Extreme Heat and Cold. *International Archives of Clinical Physiology*, 1(1), 12 2018.

[36] Malcolm N. Mistry, Rochelle Schneider, Pierre Masselot, Dominic Royé, Ben Armstrong, Jan Kyselý, Hans Orru, Francesco Sera, Shilu Tong, Éric Lavigne, Aleš Urban, Joana Madureira, David

García-León, Dolores Ibarreta, Juan Carlos Ciscar, Luc Feyen, Evan de Schrijver, Micheline de Sousa Zanotti Stagliorio Coelho, Mathilde Pascal, Aurelio Tobias, Yuming Yue Liang Leon Guo, Ana M. Vicedo-Cabrera, Antonio Gasparrini, Barrak Alahmad, Rosana Abrutzky, Paulo Hilario Nascimento Saldiva, Patricia Matus Correa, Nicolás Valdés Orteg, Haidong Kan, Samuel Osorio, Ene Indermitte, Jouni J.K. Jaakkola, Niilo Ryti, Alexandra Schneider, Veronika Huber, Klea Katsouyanni, Antonis Analitis, Alireza Entezari, Fatemeh Mayvaneh, Paola Michelozzi, Francesca de'Donato, Masahiro Hashizume, Yoonhee Kim, Magali Hurtado Diaz, César De la Cruz Valencia, Ala Overcenco, Danny Houthuijs, Caroline Ameling, Shilpa Rao, Xerxes Seposo, Baltazar Nunes, Iulian Horia Holobaca, Ho Kim, Whanhee Lee, Carmen Íñiguez, Bertil Forsberg, Christofer Åström, Martina S. Ragettli, Yuming Yue Liang Leon Guo, Bing Yu Chen, Valentina Colistro, Antonella Zanobetti, Joel Schwartz, Tran Ngoc Dang, Do Van Dung, Yuming Yue Liang Leon Guo, Ana M. Vicedo-Cabrera, and Antonio Gasparrini. Comparison of weather station and climate reanalysis data for modelling temperature-related mortality. *Scientific Reports*, 12(1):1–14, 3 2022.

[37] Evan de Schrijver, Christophe L. Folly, Rochelle Schneider, Dominic Royé, Oscar H. Franco, Antonio Gasparrini, and Ana M. Vicedo-Cabrera. A Comparative Analysis of the Temperature-Mortality Risks Using Different Weather Datasets Across Heterogeneous Regions. *GeoHealth*, 5(5):1–14, 2021.

[38] A. G. Barnett, S. Tong, and A. C.A. Clements. What measure of temperature is the best predictor of mortality? *Environmental Research*, 110(6):604–611, 8 2010.

[39] Zhiwei Xu, Gerard FitzGerald, Yuming Guo, Bin Jalaludin, and Shilu Tong. Impact of heatwave on mortality under different heatwave definitions: A systematic review and meta-analysis. *Environment international*, 89-90:193–203, 4 2016.

[40] Antonio Gasparrini, Yuming Guo, Masahiro Hashizume, Patrick L. Kinney, Elisaveta P. Petkova, Eric Lavigne, Antonella Zanobetti, Joel D. Schwartz, Aurelio Tobias, Michela Leone, Shilu Tong, Yasushi Honda, Ho Kim, and Ben G. Armstrong. Temporal variation in heat–mortality associations: A multicountry study. *Environmental Health Perspectives*, 123(11):1200–1207, 2015.

[41] Katherine Arbuthnott, Shakoor Hajat, Clare Heaviside, and Sotiris Vardoulakis. Changes in population susceptibility to heat and cold over time: assessing adaptation to climate change. *Environmental health : a global access science source*, 15 Suppl 1:33, 3 2016.

[42] Katrin G. Burkart, Michael Brauer, Aleksandr Y. Aravkin, William W. Godwin, Simon I. Hay, Jaiwei He, Vincent C. Iannucci, Samantha L. Larson, Stephen S. Lim, Jiangmei Liu, Christo-

pher J.L. Murray, Peng Zheng, Maigeng Zhou, and Jeffrey D. Stanaway. Estimating the cause-specific relative risks of non-optimal temperature on daily mortality: a two-part modelling approach applied to the Global Burden of Disease Study. *The Lancet*, 398(10301):685–697, 2021.

[43] Ana Maria Vicedo-Cabrera, Yuming Guo, Francesco Sera, Veronika Huber, Carl Friedrich Schleussner, Dann Mitchell, Shilu Tong, Micheline de Sousa Zanotti Stagliorio Coelho, Paulo Hilario Nascimento Saldiva, Eric Lavigne, Patricia Matus Correa, Nicolas Valdes Ortega, Haidong Kan, Samuel Osorio, Jan Kyselý, Aleš Urban, Jouni J.K. Jaakkola, Niilo R.I. Ryti, Mathilde Pascal, Patrick G. Goodman, Ariana Zeka, Paola Michelozzi, Matteo Scortichini, Masahiro Hashizume, Yasushi Honda, Magali Hurtado-Diaz, Julio Cruz, Xerxes Seposo, Ho Kim, Aurelio Tobias, Carmen Íñiguez, Bertil Forsberg, Daniel Oudin Åström, Martina S. Ragettli, Martin Röösli, Yue Leon Guo, Chang fu Wu, Antonella Zanobetti, Joel Schwartz, Michelle L. Bell, Tran Ngoc Dang, Dung Do Van, Clare Heaviside, Sotiris Vardoulakis, Shakoor Hajat, Andy Haines, Ben Armstrong, Kristie L. Ebi, and Antonio Gasparrini. Temperature-related mortality impacts under and beyond Paris Agreement climate change scenarios. *Climatic Change*, 150(3-4):391–402, 10 2018.

[44] R Daniel Bressler, Frances C Moore, Kevin Rennert, and David Anthoff. Estimates of country level temperature-related mortality damage functions. *Scientific Reports —*, 11:20282, 123.

[45] Antonella Zanobetti and Joel Schwartz. Temperature and mortality in nine US cities. *Epidemiology*, 19(4):563–570, 7 2008.

[46] Michela Baccini, Annibale Biggeri, Gabriele Accetta, Tom Kosatsky, Klea Katsouyanni, Antonis Analitis, H. Ross Anderson, Luigi Bisanti, Daniela D'Iippoliti, Jana Danova, Bertil Forsberg, Sylvia Medina, Anna Paldy, Daniel Rabczenko, Christian Schindler, and Paola Michelozzi. Heat effects on mortality in 15 European cities. *Epidemiology (Cambridge, Mass.)*, 19(5):711–719, 9 2008.

[47] Ru Cao, Yuxin Wang, Jing Huang, Jie He, Pitakchon Ponsawansong, Jianbo Jin, Zhihu Xu, Teng Yang, Xiaochuan Pan, Tippawan Prapamontol, and Guoxing Li. The Mortality Effect of Apparent Temperature: A Multi-City Study in Asia. *International journal of environmental research and public health*, 18(9), 2021.

[48] Susanna Conti, Paola Meli, Giada Minelli, Renata Solimini, Virgilia Toccaceli, Monica Vichi, Carmen Beltrano, and Luigi Perini. Epidemiologic study of mortality during the Summer 2003 heat wave in Italy. *Environmental Research*, 98(3):390–399, 7 2005.

[49] Keith R. Spangler, Shixin Liang, and Gregory A. Wellenius. Wet-Bulb Globe Temperature, Uni-

versal Thermal Climate Index, and Other Heat Metrics for US Counties, 2000–2020. *Scientific Data 2022 9:1*, 9(1):1–9, 6 2022.

[50] Elizabeth Bailey, Christopher Fuhrmann, Jennifer Runkle, Scott Stevens, Michael Brown, and Margaret Sugg. Wearable sensors for personal temperature exposure assessments: A comparative study. *Environmental research*, 180, 1 2020.

[51] Jennifer L. Nguyen, Joel Schwartz, and Douglas W. Dockery. The relationship between indoor and outdoor temperature, apparent temperature, relative humidity, and absolute humidity. *Indoor Air*, 24(1):103–112, 2014.

[52] Ben Armstrong, Francesco Sera, Ana Maria Vicedo-Cabrera, Rosana Abrutzky, Daniel Oudin Åström, Michelle L. Bell, Bing Yu Chen, Micheline de Sousa Zanotti Stagliorio Coelho, Patricia Matus Correa, Tran Ngoc Dang, Magali Hurtado Diaz, Do Van Dung, Bertil Forsberg, Patrick Goodman, Yue Liang Leon Guo, Yuming Guo, Masahiro Hashizume, Yasushi Honda, Ene Indermitte, Carmen Íñiguez, Haidong Kan, Ho Kim, Jan Kyselý, Eric Lavigne, Paola Michelozzi, Hans Orru, Nicolás Valdés Ortega, Mathilde Pascal, Martina S. Ragettli, Paulo Hilario Nascimento Saldiva, Joel Schwartz, Matteo Scortichini, Xerxes Seposo, Aurelio Tobias, Shilu Tong, Aleš Urban, César De la Cruz Valencia, Antonella Zanobetti, Ariana Zeka, and Antonio Gasparrini. The role of humidity in associations of high temperature with mortality: A multicountry, multicity study. *Environmental Health Perspectives*, 127(9):097007–1, 9 2019.

[53] Shakoor Hajat, Ben G. Armstrong, Nelson Gouveia, and Paul Wilkinson. Mortality displacement of heat-related deaths: A comparison of Delhi, São Paulo, and London. *Epidemiology*, 16(5):613–620, 2005.

[54] Alfésio L.F. Braga, Antonella Zanobetti, and Joel Schwartz. The effect of weather on respiratory and cardiovascular deaths in 12 U.S. cities. *Environmental Health Perspectives*, 110(9):859–863, 2002.

[55] Michael V. Saha, Robert E. Davis, and David M. Hondula. Mortality displacement as a function of heat event strength in 7 us cities. *American Journal of Epidemiology*, 179(4):467–474, 2014.

[56] Ben Armstrong, Antonio Gasparrini, and Shakoor Hajat. Estimating mortality displacement during and after heat waves. *American journal of epidemiology*, 179(12):1405–1406, 6 2014.

[57] Marshall Burke, Felipe González, Patrick Baylis, Sam Heft-Neal, Ceren Baysan, Sanjay Basu, and Solomon Hsiang. Higher temperatures increase suicide rates in the United States and Mexico. *Nature Climate Change*, 8(8):723–729, 2018.

[58] Joel Schwartz. Harvesting and long term exposure effects in the relation between air pollution and mortality. *American Journal of Epidemiology*, 151(5):440–448, 2000.

[59] Antonio Gasparrini. Modeling exposure-lag-response associations with distributed lag non-linear models. *Statistics in Medicine*, 33(5):881–899, 2014.

[60] A Gasparrini, B Armstrong, and M. G. Kenward. Multivariate meta-analysis for non-linear and other multi-parameter.pdf. *Statistics in Medicine*, 31:3821–3839, 2012.

[61] Antonio Gasparrini and Ben Armstrong. Reducing and meta-analysing estimates from distributed lag non-linear models. *BMC Medical Research Methodology*, 13(1):1–10, 2013.

[62] Maicolm Maclure. The case-crossover design: A method for studying transient effects on the risk of acute events. *American Journal of Epidemiology*, 133(2), 1991.

[63] Yao Wu, Shanshan Li, and Yuming Guo. Space-Time-Stratified Case-Crossover Design in Environmental Epidemiology Study. *Health Data Science*, 2021, 2021.

[64] TF Bateson and J Schwartz. Control for seasonal variation and time trend in case-crossover studies of acute effects of environmental exposures. *Epidemiology*, 10(4), 1999.

[65] Drew Levy, Thomas Lumley, Lianne Sheppard, Joel Kaufman, and Harvey Checkoway. Referent Selection in Case-Crossover (2001).pdf. *Epidemiology*, 12(2):186–192, 2001.

[66] Thomas F. Bateson and Joel Schwartz. Selection bias and confounding in case-crossover analyses of environmental time-series data. *Epidemiology*, 12(6):654–661, 2001.

[67] Apolline Saucy, Martina S. Ragettli, Danielle Vienneau, Kees de Hoogh, Louise Tangermann, Beat Schäffer, Jean Marc Wunderli, Nicole Probst-Hensch, and Martin Röösli. The role of extreme temperature in cause-specific acute cardiovascular mortality in Switzerland: A case-crossover study. *Science of the Total Environment*, 790:147958, 2021.

[68] Paul J. Villeneuve, David Huynh, Éric Lavigne, Ian Colman, Hymie Anisman, Cheryl Peters, and Laura A. Rodríguez-Villamizar. Daily changes in ambient air pollution concentrations and temperature and suicide mortality in Canada: Findings from a national time-stratified case-crossover study. *Environmental Research*, 223(January), 2023.

[69] Martina S. Ragettli, Apolline Saucy, Benjamin Flückiger, Danielle Vienneau, Kees de Hoogh, Ana M. Vicedo-Cabrera, Christian Schindler, and Martin Röösli. Explorative Assessment of the Temperature–Mortality Association to Support Health-Based Heat-Warning Thresholds: A National Case-Crossover Study in Switzerland. *International Journal of Environmental Research and Public Health*, 20(6):4958, 2023.

[70] Sze Hang Fu, Antonio Gasparrini, Peter S. Rodriguez, and Prabhat Jha. Mortality attributable to hot and cold ambient temperatures in India: a nationally representative case-crossover study. *PLoS Medicine*, 15(7):1–17, 2018.

[71] Shilu Tong, Xiao Yu Wang, and Yuming Guo. Assessing the short-term effects of heatwaves on mortality and morbidity in Brisbane, Australia: Comparison of case-crossover and time series analyses. *PLoS ONE*, 7(5):1–7, 2012.

[72] Ismael Henrique Silveira, Taísa Rodrigues Cortes, Beatriz Fátima Alves Oliveira, and Washington Leite Junger. Temperature and cardiovascular mortality in Rio de Janeiro, Brazil: Effect modification by individual-level and neighbourhood-level factors. *Journal of Epidemiology and Community Health*, 75(1):69–75, 2021.

[73] Joel Schwartz. How sensitive is the association between ozone and daily deaths to control for temperature? *American Journal of Respiratory and Critical Care Medicine*, 171(6):627–631, 2005.

[74] Rupa Basu, Francesca Dominici, and Jonathan M. Samet. Temperature and mortality among the elderly in the United States: A comparison of epidemiologic methods. *Epidemiology*, 16(1):58–66, 2005.

[75] Ben Armstrong. Models for the relationship between ambient temperature and daily mortality. *Epidemiology*, 17(6):624–631, 2006.

[76] A. Gasparrini, B. Armstrong, and M. G. Kenward. Distributed lag non-linear models. *Statistics in Medicine*, 29(21):2224–2234, 2010.

[77] Antonio Gasparrini and Ben Armstrong. Distributed lag non-linear models in R: the package dlnm. *Available: cran. r-project. org/web/packages/dlnm/vignettes/dlnmOverview. pdf (accessed 20 July 2011]*, 43(8):1–20, 2011.

[78] Roger D. Peng, Francesca Dominici, and Thomas A Louis. Model choice in time series studies of air pollution and mortality. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, 169(2):179–203, 2006.

[79] Peng Zhao and Zhibin He. A First Evaluation of ERA5-Land Reanalysis Temperature Product Over the Chinese Qilian Mountains. *Frontiers in Earth Science*, 10(August):1–10, 2022.

[80] Yun Zhou, Jingju Pan, Ruijun Xu, Wenfeng Lu, Yaqi Wang, Tingting Liu, Zhaoyu Fan, Yingxin Li, Chunxiang Shi, Lan Zhang, Yuewei Liu, and Hong Sun. Asthma mortality attributable to ambient temperatures: A case-crossover study in China. *Environmental Research*, 214(P4):114116, 2022.

[81] Chengcheng Qiu, John S. Ji, and Michelle L. Bell. Effect modification of greenness on temperature-mortality relationship among older adults: A case-crossover study in China. *Environmental Research*, 197(8):111112, 2021.

[82] Lidia Casas, Bianca Cox, Benoit Nemery, Patrick Deboosere, and Tim S. Nawrot. High temperatures trigger suicide mortality in Brussels, Belgium: A case-crossover study (2002–2011). *Environmental Research*, 207(October 2021):112159, 2022.

[83] Antonio Gasparrini. A tutorial on the case time series design for small-area analysis. *BMC Medical Research Methodology*, 22(1):1–8, 2022.

[84] Hicham Achebak, Daniel Devolder, and Joan Ballester. Heat-related mortality trends under recent climate warming in Spain: A 36-year observational study. *PLoS Med*, 15(7):e1002617, 7 2018.

[85] Alan Barreca, Karen Clay, Olivier Deschenes, Michael Greenstone, and Joseph S. Shapiro. Adapting to climate change: The remarkable decline in the US temperature-mortality relationship over the Twentieth Century. *Journal of Political Economy*, 124(1):105–159, 2016.

[86] Yeonseung Chung, Daewon Yang, Antonio Gasparrini, Ana M Vicedo-Cabrera, Chris Fook Sheng Ng, Yoonhee Kim, Yasushi Honda, and Masahiro Hashizume. Changing Susceptibility to Non-Optimum Temperatures in Japan, 1972-2012: The Role of Climate, Demographic, and Socioeconomic Factors. *Environmental health perspectives*, 126(5):057002, 5 2018.

[87] Seulkee Heo, Eunil Lee, Bo Yeon Kwon, Suji Lee, Kyung Hee Jo, and Jinsun Kim. Long-term changes in the heat-mortality relationship according to heterogeneous regional climate: A time-series study in South Korea. *BMJ Open*, 6:e011786, 2016.

[88] Patrick L. Kinney. Temporal Trends in Heat-Related Mortality: Implications for Future Projections. *Atmosphere 2018, Vol. 9, Page 409*, 9(10):409, 10 2018.

[89] Hicham Achebak, Daniel Devolder, and Joan Ballester. Trends in temperature-related age-specific and sex-specific mortality from cardiovascular diseases in Spain: a national time-series analysis. *The Lancet Planetary Health*, 3(7):e297–e306, 7 2019.

[90] Evan de Schrijver, Marvin Bundo, Martina S. Ragettli, Francesco Sera, Antonio Gasparrini, Oscar H. Franco, and Ana M. Vicedo-Cabrera. Nationwide Analysis of the Heat-and Cold-Related Mortality Trends in Switzerland between 1969 and 2017: The Role of Population Aging. *Environmental Health Perspectives*, 130(3):1–9, 3 2022.

[91] Scott C. Sheridan, P. Grady Dixon, Adam J. Kalkstein, and Michael J. Allen. Recent Trends in Heat-Related Mortality in the United States: An Update through 2018. *Weather, Climate, and Society*, 13(1):95–106, 1 2021.

[92] Marta Ellena, Joan Ballester, Giuseppe Costa, and Hicham Achebak. Evolution of temperature-attributable mortality trends looking at social inequalities: An observational case study of urban maladaptation to cold and heat. *Environmental Research*, 214:114082, 11 2022.

[93] Peng Lu, Jiaming Miao, Shurong Feng, al , Ji-Young Son, Jia Coco Liu, Michelle L Bell, Jinjian Chen, Maigeng Zhou, Jun Yang, Qi Zhao, Guoxin Xia, Rongbin Xu, Liz Hanna, Jing Jiang, Shanshan Li, and Yuming Guo. Temporal trends of the association between ambient temperature and cardiovascular mortality: a 17-year case-crossover study. *Environmental Research Letters*, 16(4):045004, 3 2021.

[94] Mireille A. Folkerts, Peter Bröde, W. J.Wouter Botzen, Mike L. Martinius, Nicola Gerrett, Carel N. Harmsen, and Hein A.M. Daanen. Long Term Adaptation to Heat Stress: Shifts in the Minimum Mortality Temperature in the Netherlands. *Frontiers in Physiology*, 11(March):1–7, 2020.

[95] Antonio Gasparrini, Yuming Guo, Masahiro Hashizume, Eric Lavigne, Aurelio Tobias, Antonella Zanobetti, Joel D. Schwartz, Michela Leone, Paola Michelozzi, Haidong Kan, Shilu Tong, Yasushi Honda, Ho Kim, and Ben G. Armstrong. Changes in Susceptibility to Heat during the Summer: A Multicountry Analysis. *American Journal of Epidemiology*, 183(11):1027–1036, 6 2016.

[96] Claire R. Lay, Marcus C. Sarofim, Alina Vodonos Zilberg, Dave M. Mills, Russell W. Jones, Joel Schwartz, and Patrick L. Kinney. City-level vulnerability to temperature-related mortality in the USA and future projections: a geographically clustered meta-regression. *The Lancet Planetary Health*, 5(6):e338–e346, 2021.

[97] Whanhee Lee, Hayon Michelle Choi, Jae Young Lee, Da Hye Kim, Yasushi Honda, and Ho Kim. Temporal changes in mortality impacts of heat wave and cold spell in Korea and Japan. *Environment International*, 116(December 2017):136–146, 2018.

[98] Keith R. Spangler and Gregory A. Wellenius. Spatial and intraseasonal variation in changing susceptibility to extreme heat in the United States. *Environmental Epidemiology*, 5(2):e136, 2021.

[99] Shengzhi Sun, Wangnan Cao, Tonya G Mason, Jinjun Ran, Hong Qiu, Jinhui Li, Yang Yang,

Hualiang Lin, and Linwei Tian. Increased susceptibility to heat for respiratory hospitalizations in Hong Kong. *The Science of the total environment*, 666:197–204, 5 2019.

[100] Aurelio Tobías, Masahiro Hashizume, Yasushi Honda, Francesco Sera, Chris Fook Sheng Ng, Yoonhee Kim, Dominic Roye, Yeonseung Chung, Tran Ngoc Dang, Ho Kim, Whanhee Lee, Carmen Íñiguez, Ana Vicedo-Cabrera, Rosana Abrutzky, Yuming Guo, Shilu Tong, Micheline de Sousa Zanotti Stagliorio Coelho, Paulo Hilario Nascimento Saldiva, Eric Lavigne, Patricia Matus Correa, Nicolás Valdés Ortega, Haidong Kan, Samuel Osorio, Jan Kyselý, Aleš Urban, Hans Orru, Ene Indermitte, Jouni J. K. Jaakkola, Niilo R. I. Ryti, Mathilde Pascal, Veronika Huber, Alexandra Schneider, Klea Katsouyanni, Antonis Analitis, Alireza Entezari, Fatemeh Mayvaneh, Patrick Goodman, Ariana Zeka, Paola Michelozzi, Francesca de'Donato, Barrak Alahmad, Magali Hurtado Diaz, César De la Cruz Valencia, Ala Overcenco, Danny Houthuijs, Caroline Ameling, Shilpa Rao, Francesco Di Ruscio, Gabriel Carrasco, Xerxes Seposo, Baltazar Nunes, Joana Madureira, Iulian-Horia Holobaca, Noah Scovronick, Fiorella Acquaotta, Bertil Forsberg, Christofer Åström, Martina S. Ragettli, Yue-Liang Leon Guo, Bing-Yu Chen, Shanshan Li, Valentina Colistro, Antonella Zanobetti, Joel Schwartz, Do Van Dung, Ben Armstrong, and Antonio Gasparrini. Geographical Variations of the Minimum Mortality Temperature at a Global Scale. *Environmental Epidemiology*, 5(5):e169, 2021.

[101] Lisbeth Weitensfelder and Hanns Moshammer. Evidence of Adaptation to Increasing Temperatures. pages 1971–1975, 2020.

[102] Daniel Oudin Åström, Kristie L. Ebi, Ana Maria Vicedo-Cabrera, and Antonio Gasparrini. Investigating changes in mortality attributable to heat and cold in Stockholm, Sweden. *International Journal of Biometeorology*, 62(9):1777–1780, 2018.

[103] Ana M. Vicedo-Cabreraa, Francesco Sera, Yuming Guo, Yeonseung Chung, Katherine Arbuthnott, Shilu Tong, Aurelio Tobias, Eric Lavigne, Micheline de Sousa Zanotti Stagliorio Coelho, Paulo Hilario Nascimento Saldiva, Patrick G. Goodman, Ariana Zeka, Masahiro Hashizume, Yasushi Honda, Ho Kim, Martina S. Ragettli, Martin Röösli, Antonella Zanobetti, Joel Schwartz, Ben Armstrong, and Antonio Gasparrinio. A multi-country analysis on potential adaptive mechanisms to cold and heat in a changing climate. *Environment International*, 111:239–246, 2 2017.

[104] Camilo Mora, Bénédicte Dousset, Iain R. Caldwell, Farrah E. Powell, Rollan C. Geronimo, Coral R. Bielecki, Chelsie W.W. Counsell, Bonnie S. Dietrich, Emily T. Johnston, Leo V. Louis, Matthew P. Lucas, Marie M. Mckenzie, Alessandra G. Shea, Han Tseng, Thomas W. Giambelluca, Lisa R. Leon, Ed Hawkins, and Clay Trauernicht. Global risk of deadly heat. *Nature Climate Change*, 7(7):501–506, 2017.

[105] Jingwen Liu, Alana Hansen, Blesson Varghese, Zhidong Liu, Michael Tong, Hong Qiu, Linwei Tian, Kevin Ka Lun Lau, Edward Ng, Chao Ren, and Peng Bi. Cause-specific mortality attributable to cold and hot ambient temperatures in Hong Kong: a time-series study, 2006–2016. *Sustainable Cities and Society*, 57(March):102131, 2020.

[106] Josiah L. Kephart, Brisa N. Sánchez, Jeffrey Moore, Leah H. Schinasi, Maryia Bakhtsiyarava, Yang Ju, Nelson Gouveia, Waleska T. Caiaffa, Iryna Dronova, Saravanan Arunachalam, Ana V. Diez Roux, and Daniel A. Rodríguez. City-level impact of extreme temperatures and mortality in Latin America. *Nature Medicine 2022 28:8*, 28(8):1700–1705, 6 2022.

[107] Sara Lopes de Moraes, Ricardo Almendra, and Ligia Vizeu Barrozo. Impact of heat waves and cold spells on cause-specific mortality in the city of São Paulo, Brazil. *International Journal of Hygiene and Environmental Health*, 239:1–28, 1 2022.

[108] Ji Young Son, Nelson Gouveia, Mercedes A. Bravo, Clarice Umbelino de Freitas, and Michelle L. Bell. The impact of temperature on mortality in a subtropical city: effects of cold, heat, and heat waves in São Paulo, Brazil. *International Journal of Biometeorology*, 60(1):113–121, 2016.

[109] Itai Kloog, Alexandra Chudnovsky, Petros Koutrakis, and Joel Schwartz. Temporal and spatial assessments of minimum air temperature using satellite surface temperature measurements in Massachusetts, USA. *Science of the Total Environment*, 432:85–92, 2012.

[110] Renjie Chen, Peng Yin, Lijun Wang, Cong Liu, Yue Niu, Weidong Wang, Yixuan Jiang, Yunning Liu, Jiangmei Liu, Jinlei Qi, Jinling You, Haidong Kan, and Maigeng Zhou. Association between ambient temperature and mortality risk and burden: time series study in 272 main Chinese cities. *BMJ (Clinical research ed.)*, 363:k4306, 10 2018.

[111] Rochelle Schneider dos Santos. Estimating spatio-temporal air temperature in London (UK) using machine learning and earth observation satellite data. *International Journal of Applied Earth Observation and Geoinformation*, 88(December 2019), 2020.

[112] Hendrik Wouters, Koen De Ridder, Lien Poelmans, Patrick Willems, Johan Brouwers, Parisa Hosseinzadehtalaei, Hossein Tabari, Sam Vanden Broucke, Nicole P.M. van Lipzig, and Matthias Demuzere. Heat stress increase under climate change twice as large in cities as in rural areas: A study for a densely populated midlatitude maritime region. *Geophysical Research Letters*, 44(17):8997–9007, 2017.

[113] Jack J. Lennon and John R.G. Turner. Predicting the Spatial Distribution of Climate: Temperature in Great Britain. *The Journal of Animal Ecology*, 64(3):370, 1995.

[114] David J. Mildrexler, Maosheng Zhao, and Steven W. Running. A global comparison between station air temperatures and MODIS land surface temperatures reveals the cooling role of forests. *Journal of Geophysical Research: Biogeosciences*, 116(3):1–15, 2011.

[115] European Centre for Medium-Range Weather Forecasts. Datasets — ECMWF, 2020.

[116] Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, Adrian Simmons, Cornel Soci, Saleh Abdalla, Xavier Abellan, Gianpaolo Balsamo, Peter Bechtold, Gionata Biavati, Jean Bidlot, Massimo Bonavita, Giovanna De Chiara, Per Dahlgren, Dick Dee, Michail Diamantakis, Rossana Dragani, Johannes Flemming, Richard Forbes, Manuel Fuentes, Alan Geer, Leo Haimberger, Sean Healy, Robin J. Hogan, Elías Hólm, Marta Janisková, Sarah Keeley, Patrick Laloyaux, Philippe Lopez, Cristina Lupu, Gabor Radnoti, Patricia de Rosnay, Iryna Rozum, Freja Vamborg, Sebastien Villaume, and Jean Noël Thépaut. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, 7 2020.

[117] Joaquín Muñoz-Sabater, Emanuel Dutra, Anna Agustí-Panareda, Clément Albergel, Gabriele Arduini, Gianpaolo Balsamo, Souhail Boussetta, Margarita Choulga, Shaun Harrigan, Hans Hersbach, Brecht Martens, Diego G. Miralles, Mariá Piles, Nemesio J. Rodríguez-Fernández, Ervin Zsoter, Carlo Buontempo, and Jean Noël Thépaut. ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, 13(9):4349–4383, 9 2021.

[118] S. Schimanke, M. Ridal, P. Le Moigne, L. Berggren, P. Undén, R. Randriamampianina, U. Andrea, E. Bazile, A. Bertelsen, P. Brousseau, P. Dahlgren, L. Edvinsson, A. El Said, M. Glinton, S. Hopsch, L. Isaksson, R. Mladek, E. Olsson, A. Verrelle, and Z.Q. Wang. CERRA sub-daily regional reanalysis data for Europe on single levels from 1984 to present. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*, 2021.

[119] WR. Tobler. A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46(Supplement: Proceedings. International Geographical Union. Commission on Quantitative Methods (Jun., 1970)):234–240, 1970.

[120] David F Watson. A refinement of inverse distance weighted interpolation. *Geo-processing*, 2:315–327, 1985.

[121] Jianbo Tan, Xinyao Xie, Jiaqi Zuo, Xuemin Xing, Bin Liu, Qing Xia, and Yunfei Zhang. Coupling random forest and inverse distance weighting to generate climate surfaces of precipitation and temperature with Multiple-Covariates. *Journal of Hydrology*, 598, 2021.

[122] Carlotta Quagliolo, Alessandro Pezzoli, Rosaria Ignaccolo, and José Luis Santos Davila. Time-lagged inverse-distance weighting for air temperature analysis in an equatorial urban area (Guayaquil, Ecuador). *Meteorological Applications*, 27(4):1–16, 2020.

[123] Aleksandar Sekulić, Milan Kilibarda, Gerard B.M. Heuvelink, Mladen Nikolić, and Branislav Bajat. Random forest spatial interpolation. *Remote Sensing*, 12(10):1–29, 2020.

[124] Milan Kilibarda, Tomislav Hengl, Gerard B.M. Heuvelink, Benedikt Gräler, Edzermelita Perčec Pebesmatadić, and Branislav Bajat. Spatio-temporal interpolation of daily temperatures for global land areas at 1 km resolution. *Journal of Geophysical Research*, 119(5):2294–2313, 2014.

[125] Dominique Courault and Pascal Monestiez. Spatial interpolation of air temperature according to atmospheric circulation patterns in southeast France. *International Journal of Climatology*, 19(4):365–378, 1999.

[126] Christopher K. Wikle, Andrew Zammit-Mangion, and Noel Cressie. *Spatio-Temporal Statistics With R*. Chapman & Hall/CRC, 2019.

[127] Richard D. Hunter and Ross K. Meentemeyer. Climatologically aided mapping of daily precipitation and temperature. *Journal of Applied Meteorology*, 44(10):1501–1510, 2005.

[128] J. J. Carrera-Hernández and S. J. Gaskin. Spatio temporal analysis of daily precipitation and temperature in the Basin of Mexico. *Journal of Hydrology*, 336(3-4):231–249, 2007.

[129] Nynke Hofstra, Malcolm Haylock, Mark New, Phil Jones, and Christoph Frei. Comparison of six methods for the interpolation of daily, European climate data. *Journal of Geophysical Research Atmospheres*, 113(21), 2008.

[130] Edzer J. Pebesma. Multivariable geostatistics in S: The gstat package. *Computers and Geosciences*, 30(7):683–691, 2004.

[131] Benedikt Gräler, Edzer Pebesma, and Gerard Heuvelink. Spatio-temporal interpolation using gstat. *R Journal*, 8(1):204–218, 2016.

[132] Leo Breiman. Random Forest. *Machine Learning*, 45:5–32, 2001.

[133] Tyler J VanderWeele. On the distinction between interaction and effect modification. *Epidemiology (Cambridge, Mass.)*, 20(6):863–871, 11 2009.

[134] Xuping Song, Shigong Wang, Yuling Hu, Man Yue, Tingting Zhang, Yu Liu, Jinhui Tian, and Kezheng Shang. Impact of ambient temperature on morbidity and mortality: An overview of reviews. *Science of the Total Environment*, 586(222):241–254, 2017.

[135] Tarik Benmarhnia, Severine Deguen, Jay S Kaufman, and Audrey Smargiassi. Review Article: Vulnerability to Heat-related Mortality: A Systematic Review, Meta-analysis, and Meta-regression Analysis. *Epidemiology (Cambridge, Mass.)*, 26(6):781–793, 11 2015.

[136] Shakoor Hajat, Scott C. Sheridan, Michael J. Allen, Mathilde Pascal, Karine Laaidi, Abderrahmane Yagouti, Ugis Bickis, Aurelio Tobias, Denis Bourque, Ben G. Armstrong, and Tom Kosatsky. Heat-health warning systems: A comparison of the predictive capacity of different approaches to identifying dangerously hot days. *American Journal of Public Health*, 100(6):1137–1144, 2010.

[137] Alistair Woodward. Heat, cold and climate change. *Journal of Epidemiology and Community Health*, 68(7):595–596, 2014.

[138] Massimo Stafoggia, Francesco Forastiere, Daniele Agostini, Annibale Biggeri, Luigi Bisanti, Ennio Cadum, Nicola Caranci, Francesco De'Donato, Sara De Lisio, Moreno De Maria, Paola Michelozzi, Rossella Miglio, Paolo Pandolfi, Sally Picciotto, Magda Rognoni, Antonio Russo, Corrado Scarnato, and Carlo A. Perucci. Vulnerability to heat-related mortality: A multicity, population-based, case-crossover analysis. *Epidemiology*, 17(3):315–323, 2006.

[139] Rupa Basu. High ambient temperature and mortality: a review of epidemiologic studies from 2001 to 2008. *Environmental health : a global access science source*, 8:40, 9 2009.

[140] Antonella Zanobetti, Marie S O'Neill, Carina J Gronlund, and Joel D Schwartz. Susceptibility to mortality in weather extremes: effect modification by personal and small-area characteristics. *Epidemiology (Cambridge, Mass.)*, 24(6):809–819, 11 2013.

[141] Jun Yang, Chun-Quan Ou, Yan Ding, Ying-Xue Zhou, and Ping-Yan Chen. Daily temperature and mortality: a study of distributed lag non-linear effect and effect modification in Guangzhou. *Environmental health : a global access science source*, 11:63, 9 2012.

[142] Alessandro Marinaccio, Matteo Scortichini, Claudio Gariazzo, Antonio Leva, Michela Bonafede, Francesca K. de' Donato, Massimo Stafoggia, Giovanni Viegi, Paola Michelozzi, Ancona Carla, Angelini Paola, Argentini Stefania, Baldacci Sandra, Bisceglia Lucia, Bonomo Sergio, Bonvicini Laura, Broccoli Serena, Brusasca Giuseppe, Bucci Simone, Calori Giuseppe, Carlino Giuseppe, Achille Cernigliaro, Chieti Antonio, Fasola Salvatore, Finardi Sandro, Forastiere Francesco, Galassi Claudia, Giorgi Rossi Paolo, La Grutta Stefania, Licitra Gaetano, Maio Sara, Migliore Enrica, Moro Antonino, Nanni Alessandro, Ottone Marta, Pepe Nicola, Radice Paola, Ranzi Andrea, Renzi Matteo, Scondotto Salvatore, Silibello Camillo, Sozzi Roberto, Tinarelli Gianni, and Uboldi Francesco. Nationwide epidemiological study for estimating the effect of extreme outdoor

temperature on occupational injuries in Italy. *Environment International*, 133(May):105176, 2019.

[143] Jing Li, Xin Xu, Jun Yang, Zhidong Liu, Lei Xu, Jinghong Gao, Xiaobo Liu, Haixia Wu, Jun Wang, Jieqiong Yu, Baofa Jiang, and Qiyong Liu. Ambient high temperature and mortality in Jinan, China: A study of heat thresholds and vulnerable populations. *Environmental research*, 156:657–664, 7 2017.

[144] Jie Zeng, Xuehai Zhang, Jun Yang, Junzhe Bao, Hao Xiang, Keith Dear, Qiyong Liu, Shao Lin, Wayne R. Lawrence, Aihua Lin, and Cunrui Huang. Humidity may modify the relationship between temperature and cardiovascular mortality in Zhejiang province, China. *International Journal of Environmental Research and Public Health*, 14(11), 2017.

[145] Marie S. O'Neill, Antonella Zanobetti, and Joel Schwartz. Modifiers of the temperature and mortality association in seven US cities. *American Journal of Epidemiology*, 157(12):1074–1082, 2003.

[146] Mercedes Medina-Ramón, Antonella Zanobetti, David Paul Cavanagh, and Joel Schwartz. Extreme temperatures and mortality: Assessing effect modification by personal characteristics and specific cause of death in a multi-city case-only analysis. *Environmental Health Perspectives*, 114(9):1331–1336, 2006.

[147] Ji-Young Son, Kevin Lane, Jong-Tae Lee, and Michelle Bell. Vulnerability to temperature-related mortality in Seoul, Korea. *Environ Res Lett.*, 6(3):034027, 2011.

[148] Carme Borrell, Marc Mari-Dell'Olmo, Maica Rodriguez-Sanz, Patricia Garcia-Olalla, Joan A Cayla, Joan Benach, and Carles Muntaner. Socioeconomic position and excess mortality during the heat wave of 2003 in Barcelona. *European journal of epidemiology*, 21(9):633–640, 2006.

[149] Carina J Gronlund. Racial and socioeconomic disparities in heat-related health effects and their mechanisms: a review. *Current epidemiology reports*, 1(3):165–173, 9 2014.

[150] Alana Hansen, Linda Bi, Arthur Saniotis, and Monika Nitschke. Vulnerability to extreme heat and climate change: is ethnicity a factor? *Global health action*, 6:21364, 7 2013.

[151] S. Vandentorren, P. Bretin, A. Zeghnoun, L. Mandereau-Bruno, A. Croisier, C. Cochet, J. Ribéron, I. Siberan, B. Declercq, and M. Ledrans. August 2003 heat wave in France: Risk factors for death of elderly people living at home. *European Journal of Public Health*, 16(6):583–591, 2006.

[152] M. Stafoggia, Francesco Forastiere, D. Agostini, N. Caranci, F. De'Donato, M. Demaria, P. Michelozzi, R. Miglio, M. Rognoni, A. Russo, and C. A. Perucci. Factors affecting in-hospital heat-related mortality: A multi-city case-crossover analysis. *Journal of Epidemiology and Community Health*, 62(3):209–215, 2008.

[153] Claire Ferron, David Trewick, Philippe Le Conte, Eacute Ric Batard, Laëtitia Girard, and Gilles Potel. Heat stroke in hospital patients during the summer 2003 heat wave: a nosocomial disease. *Presse medicale (Paris, France : 1983)*, 35(2 Pt 1):196–199, 2 2006.

[154] Glen P Kenny, Andreas D Flouris, Abderrahmane Yagouti, and Sean R Notley. Towards establishing evidence-based guidelines on maximum indoor temperatures during hot weather in temperate continental climates. *Temperature (Austin, Tex.)*, 6(1):11–36, 2019.

[155] Eric Klinenberg. *Heat Wave: A Social Autopsy of Disaster in Chicago*. University of Chicago Press, Chicago, 5.

[156] Michael Marmot. Fair society, healthy lives: The Marmot Review. Technical report.

[157] Marc Marí-Dell'Olmo, Ana M. Novoa, Lluís Camprubí, Andrés Peralta, Hugo Vásquez-Vera, Jordi Bosch, Jordi Amat, Fernando Díaz, Laia Palència, Roshanak Mehdipanah, Maica Rodríguez-Sanz, Davide Malmusi, and Carme Borrell. Housing Policies and Health Inequalities. *International Journal of Health Services*, 47(2):207–232, 2017.

[158] Jonathon Taylor, Paul Wilkinson, Mike Davies, Ben Armstrong, Zaid Chalabi, Anna Mavrogianni, Phil Symonds, Eleni Oikonomou, and Sylvia I. Bohnenstengel. Mapping the effects of urban heat island, housing, and age on excess heat-related mortality in London. *Urban Climate*, 14:517–528, 2015.

[159] Francesco Sera, Ben Armstrong, Aurelio Tobias, Ana Maria Vicedo-Cabrera, Christofer Astrom, Michelle L Bell, Bing-Yu Chen, Micheline de Sousa Zanotti Stagliorio Coelho, Patricia Matus Correa, Julio Cesar Cruz, Tran Ngoc Dang, Magali Hurtado-Diaz, Dung Do Van, Bertil Forsberg, Yue Leon Guo, Yuming Guo, Masahiro Hashizume, Yasushi Honda, Carmen Iniguez, Jouni J K Jaakkola, Haidong Kan, Ho Kim, Eric Lavigne, Paola Michelozzi, Nicolas Valdes Ortega, Samuel Osorio, Mathilde Pascal, Martina S Ragettli, Niilo R I Ryti, Paulo Hilario Nascimento Saldiva, Joel Schwartz, Matteo Scortichini, Xerxes Seposo, Shilu Tong, Antonella Zanobetti, and Antonio Gasparrini. How urban characteristics affect vulnerability to heat and cold: a multi-country analysis. *International journal of epidemiology*, 48(4):1101–1112, 8 2019.

[160] Maryia Bakhtsiyarava, Leah H. Schinasi, Brisa N. Sánchez, Iryna Dronova, Josiah L. Kephart, Yang Ju, Nelson Gouveia, Waleska Teixeira Caiaffa, Marie S. O'Neill, Goro Yamada, Sarav Arunachalam, Ana V. Diez-Roux, and Daniel A. Rodríguez. Modification of temperature-related human mortality by area-level socioeconomic and demographic characteristics in Latin American cities. *Social Science and Medicine*, 317(August 2022), 2023.

[161] Antonio Gasparrini, Ben Armstrong, Sari Kovats, and Paul Wilkinson. The effect of high temperatures on cause-specific mortality in England and Wales. *Occupational and Environmental Medicine*, 69(1):56–61, 2012.

[162] Thammanitchpol Denpetkul and Arthit Phosri. Daily ambient temperature and mortality in Thailand: Estimated effects, attributable risks, and effect modifications by greenness. *Science of the Total Environment*, 791:148373, 2021.

[163] J. M. Sobstyl, T. Emig, M. J.Abdolhosseini Qomi, F. J. Ulm, and R. J.M. Pellenq. Role of City Texture in Urban Heat Islands at Nighttime. *Physical Review Letters*, 120(10):108701, 2018.

[164] Hua Shi, George Xian, Roger Auch, Kevin Gallo, and Qiang Zhou. Urban Heat Island and Its Regional Impacts Using Remotely Sensed Thermal Data—A Review of Recent Developments and Methodology. *Land*, 10(8):867, 8 2021.

[165] Cheng Li, Jie Zhao, Nguyen Xuan Thinh, Wenfu Yang, and Zhen Li. Analysis of the spatiotemporally varying effects of urban spatial patterns on land surface temperatures. *Journal of Environmental Engineering and Landscape Management*, 26(3):216–231, 2018.

[166] Benjamin F. Zaitchik and Cascade Tuholske. Earth observations of extreme heat events: Leveraging current capabilities to enhance heat research and action. *Environmental Research Letters*, 16(11), 2021.

[167] Home - MCC Collaborative Research Network.

[168] Yuming Guo, Adrian G. Barnett, Xiaochuan Pan, Weiwei Yu, and Shilu Tong. The impact of temperature on mortality in Tianjin, china: A case-crossover design with a distributed lag nonlinear model. *Environmental Health Perspectives*, 119(12):1719–1725, 2011.

[169] D. Alex Quistberg, Ana V. Diez Roux, Usama Bilal, Kari Moore, Ana Ortigoza, Daniel A. Rodriguez, Olga L. Sarmiento, Patricia Frenz, Amélia Augusta Friche, Waleska Teixeira Caiaffa, Alejandra Vives, J. Jaime Miranda, Marcio Alazraqui, Hugo Spinelli, Carlos Guevel, Vanessa Di Cecco, Adela Tisnés, Carlos Leveau, Adrián Santoro, Damián Herkovits, Nelson Gouveia, Mauricio Barreto, Gervásio Santos, Leticia Cardoso, Mariana Carvalho de Menezes,

Maria de Fatima de Pina, Amélia Augusta de Lima Friche, Amanda Cristina de Souza Andrade, Tania Alfaro, Cynthia Córdova, Pablo Ruiz, Mauricio Fuentes, Alejandra Vives Vergara, Alejandro Salazar, Andrea Cortinez-O'ryan, Cristián Schmitt, Francisca Gonzalez, Fernando Baeza, Flavia Angelini, Olga Lucía Sarmiento Dueñas, Diana Higuera, Catalina González, Felipe Montes, Andres F. Useche, Oscar Guaje, Ana Maria Jaramillo, Luis Angel Guzmán, Philipp Hessel, Diego Lucumi, Jose David Meisel, Eliana Martinez, María F. Kroker-Lobos, Manuel Ramirez-Zea, Kevin Martinez Folger, Tonatiuh Barrientos-Gutierrez, Carolina Perez-Ferrer, Javier Prado-Galbarro, Filipa de Castro, Rosalba Rojas-Martínez, Akram Hernández-Vásquez, Francisco Diez-Canseco, Ross Hammond, Iryna Dronova, Brisa N. Sanchez, Peter Hovmand, Ricardo Jordán Fuchs, Juliet Braslow, Jose Siri, Amy Auchincloss, Brent Langellier, Gina Lovasi, Leslie McClure, Yvonne Michael, Harrison Quick, Jose Tapia Granados, Felipe Garcia-España, and Ivana Stankov. Building a Data Platform for Cross-Country Urban Health Studies: the SALURBAL Study. *Journal of Urban Health*, 96(2):311–337, 4 2019.

[170] The Salud Urbana en América Latina (SALURBAL) Project: Studying Urban Health in Latin America, — Urban Health Network for Latin America and the Caribean.

[171] Josiah L. Kephart, Brisa N Sanchez, J Moore, LH Schinasi, Maryia Bakhtsiyarava, Y Ju, Nelson Gouveia, Waleska T. Caiaffa, Iryna Dronova, Saravanan Arunachalam, Ana V. Diez Roux, and Daniel A. Rodríguez. Extreme temperatures and mortality in 326 Latin American cities. *medRxiv*, pages 1–20, 2021.

[172] Maryia Bakhtsiyarava, Ana Ortigoza, Brisa N. Sánchez, Ariela Braverman-Bronstein, Josiah L. Kephart, Santiago Rodríguez López, Jordan Rodríguez, and Ana V. Diez Roux. Ambient temperature and term birthweight in Latin American cities. *Environment International*, 167(March), 2022.

[173] James E. Bennett, Marta Blangiardo, Daniela Fecht, Paul Elliott, and Majid Ezzati. Vulnerability to the mortality effects of warm temperature in the districts of England and Wales. *Nature Climate Change*, 4(4):269–273, 2014.

[174] Nicolas Todd and Alain Jacques Valleron. Space–time covariation of mortality with temperature: A systematic study of deaths in France, 1968–2009. *Environmental Health Perspectives*, 123(7):659–664, 2015.

[175] Sofia Zafeiratou, Antonis Analitis, Dimitra Founda, Christos Giannakopoulos, Konstantinos V. Varotsos, Panagiotis Sismanidis, Iphigenia Keramitsoglou, and Klea Katsouyanni. Spatial variability in the effect of high ambient temperature on mortality: An analysis at municipality level

within the greater athens area. *International Journal of Environmental Research and Public Health*, 16(19):1–20, 2019.

[176] Zhengjing Huang, Hualiang Lin, Yunning Liu, Maigeng Zhou, Tao Liu, Jianpeng Xiao, Weilin Zeng, Xing Li, Yonghui Zhang, Kristie L. Ebi, Shilu Tong, Wenjun Ma, and Lijun Wang. Individual-level and community-level effect modifiers of the temperature-mortality relationship in 66 Chinese communities. *BMJ Open*, 5(9):1–8, 2015.

[177] Mihye Lee, Liuhua Shi, Antonella Zanobetti, and Joel D. Schwartz. Study on the association between ambient temperature and mortality using spatially resolved exposure data. *Environmental Research*, 151:610–617, 2016.

[178] Marc Marí-Dell'Olmo, Aurelio Tobías, Anna Gómez-Gutiérrez, Maica Rodríguez-Sanz, Patricia García de Olalla, Esteve Camprubí, Antonio Gasparrini, and Carme Borrell. Social inequalities in the association between temperature and mortality in a South European context. *International Journal of Public Health*, 64(1):27–37, 2019.

[179] Nelson Gouveia, Shakoor Hajat, and Ben Armstrong. Socioeconomic differentials in the temperature-mortality relationship in Sao Paulo, Brazil. *International journal of epidemiology*, 32(3):390–397, 6 2003.

[180] Ghasem Sam Toloo, Gerard Fitzgerald, Peter Aitken, Kenneth Verrall, and Shilu Tong. Are heat warning systems effective? *Environmental health : a global access science source*, 12:27, 4 2013.

[181] Rachel Lowe, Markel García-Díez, Joan Ballester, James Creswick, Jean-Marie Marie Robine, François R. Francois R Herrmann, Xavier Rodó, Markel Garcia-Diez, Joan Ballester, James Creswick, Jean-Marie Marie Robine, François R. Francois R Herrmann, and Xavier Rodo. Evaluation of an Early-Warning System for Heat Wave-Related Mortality in Europe: Implications for Sub-seasonal to Seasonal Forecasting and Climate Services. *International Journal of Environmental Research and Public Health*, 13(2):206, 2 2016.

[182] Florian Pappenberger, Hannah L. Cloke, Dennis J. Parker, Fredrik Wetterhall, David S. Richardson, and Jutta Thielen. The monetary benefit of early flood warnings in Europe. *Environmental Science and Policy*, 51:278–291, 2015.

[183] Ana Casanueva, Annkatrin Burgstall, Sven Kotlarski, Alessandro Messeri, Marco Morabito, Andreas D Flouris, Lars Nybo, Christoph Spirig, and Cornelia Schwierz. Overview of Existing Heat-Health Warning Systems in Europe. *International journal of environmental research and public health*, 16(15), 7 2019.

[184] Harriet Bulkeley. Cities and the governing of climate change. *Annual Review of Environment and Resources*, 35:229–253, 2010.

[185] Mahamat Abdelkerim Issa, Fateh Chebana, Pierre Masselot, Céline Campagna, Éric Lavigne, Pierre Gosselin, and Taha B.M.J. Ouarda. A heat-health watch and warning system with extended season and evolving thresholds. *BMC Public Health*, 21(1):1–13, 2021.

[186] Tiantian Li, Chen Chen, and Wenjia Cai. The global need for smart heat–health warning systems. *The Lancet*, 400(10362):1511–1512, 2022.

[187] Kathleen E. McLean, Rebecca Stranberg, Melissa Macdonald, Gregory R.A. Richardson, Tom Kosatsky, and Sarah B. Henderson. Establishing heat alert thresholds for the varied climatic regions of British Columbia, Canada. *International Journal of Environmental Research and Public Health*, 15(9):1–14, 2018.

[188] Laurence S. Kalkstein, Paul F. Jamason, J. Scott Greene, Jerry Libby, and Lawrence Robinson. The Philadelphia hot weather-health watch/warning system: Development and application, summer 1995. *Bulletin of the American Meteorological Society*, 77(7):1519–1528, 1996.

[189] Neville Nicholls, Carol Skinner, Margaret Loughnan, and Nigel Tapper. A simple heat alert system for Melbourne, Australia. *International Journal of Biometeorology*, 52(5):375–384, 2008.

[190] World Health Organization (WHO). Health and climate change: country profile 2015: Brazil. 2015.

[191] Gabriela Marques Di Giulio, Ana Maria Barbieri Bedran-Martins, Maria da Penha Vasconcellos, Wagner Costa Ribeiro, and Maria Carmen Lemos. Mainstreaming climate adaptation in the megacity of São Paulo, Brazil. *Cities*, 72(September 2017):237–244, 2018.

[192] CEMADEN — Centro Nacional de Monitoramento e Alertas de Desastres Naturais - Cemaden/MCTI, 2023.

[193] Hylke E. Beck, Niklaus E. Zimmermann, Tim R. McVicar, Noemi Vergopolan, Alexis Berg, and Eric F. Wood. Present and future köppen-geiger climate classification maps at 1-km resolution. *Scientific Data*, 5(1):1–12, 10 2018.

[194] MeteoBlue. Climate Change São Paulo.

[195] João L. Geirinhas, Ricardo M. Trigo, Renata Libonati, Caio A.S. Coelho, and Ana Cristina Palmeira. Climatic and synoptic characterization of heat waves in Brazil. *International Journal of Climatology*, 38(4):1760–1776, 3 2018.

[196] Jhonathan F. T. de Souza1 Jacyra Soares2 Amauri Pereira de Oliveira2 Regina Maura de Miranda Flavia N. D. Ribeiro1, Arissa S. umezaki1. Urban heat island in the metropolitan area of São Paulo and the influence of warm and dry air masses during summer. 2014, 2015.

[197] Gustavo Armani, Nadia Gilma Beserra de Lima, Maria Fernanda Pelizon Garcia, and Jussara de Lima Carvalho. Regional Climate Projections for the State of São Paulo, Brazil, in the 2020 – 2050 Period. *Derbyana*, 43:1–29, 2022.

[198] Antonio Gasparrini, Yuming Guo, Francesco Sera, Ana Maria Vicedo-Cabrera, Veronika Huber, Shilu Tong, Micheline de Sousa Zanotti Stagliorio Coelho, Paulo Hilario Nascimento Saldiva, Eric Lavigne, Patricia Matus Correa, Nicolas Valdes Ortega, Haidong Kan, Samuel Osorio, Jan Kyselý, Aleš Urban, Jouni J.K. Jaakkola, Niilo R.I. Ryti, Mathilde Pascal, Patrick G. Goodman, Ariana Zeka, Paola Michelozzi, Matteo Scortichini, Masahiro Hashizume, Yasushi Honda, Magali Hurtado-Diaz, Julio Cesar Cruz, Xerxes Seposo, Ho Kim, Aurelio Tobias, Carmen Iñiguez, Bertil Forsberg, Daniel Oudin Åström, Martina S. Ragettli, Yue Leon Guo, Chang fu Wu, Antonella Zanobetti, Joel Schwartz, Michelle L. Bell, Tran Ngoc Dang, Dung Do Van, Clare Heaviside, Sotiris Vardoulakis, Shakoor Hajat, Andy Haines, and Ben Armstrong. Projections of temperature-related excess mortality under climate change scenarios. *The Lancet Planetary Health*, 1(9):e360–e367, 2017.

[199] S. Feron, R. R. Cordero, A. Damiani, P. J. Llanillo, J. Jorquera, E. Sepulveda, V. Asencio, D. Laroze, F. Labbe, J. Carrasco, and G. Torres. Observations and Projections of Heat Waves in South America. *Scientific Reports*, 9(1):1–15, 2019.

[200] Vanderli Custódio. *Escassez de água e inundações na Região Metropolitana de São Paulo*. Humanitas/ FAPESP, 2012.

[201] Segundo Relatório, Eduardo Marques, Mariana Bittar, Donizete Cazolato, Edgard Fusaro, and Daniel Waldvogel. Diagnóstico dos assentamentos precários nos municípios da Macrometrópole Paulista. 2013.

[202] Gabinete do Prefeito - Governo do Estado de São Paulo. Lei 16.050 de 31 de julho de 2014: Política de Desenvolvimento Urbano e o Plano Diretor Estratégico do Município de São Paulo, 2014.

[203] Fernanda Rodrigues Diniz, Fábio Luiz Teixeira Gonçalves, and Scott Sheridan. Heat wave and elderly mortality: Historical analysis and future projection for metropolitan region of Sao Paulo, Brazil. *Atmosphere*, 11(9), 2020.

[204] Eliane Simões, Wilson C. de Sousa Junior, Débora M. de Freitas, Morena Mills, Allan Y. Iwama, Isabel Gonçalves, Débora Olivato, and Pedro Fidelman. Barriers and opportunities for adapting to climate change on the North Coast of São Paulo, Brazil. *Regional Environmental Change*, 17(6):1739–1750, 2017.

[205] Gard Lindseth. The Cities for Climate Protection Campaign (CCPC) and the framing of Local Climate Policy. *Local Environment*, 9(4):325–336, 2004.

[206] C40 Cities - A global network of mayors taking urgent climate action.

[207] United Nations. Race to Zero.

[208] United Nations. Race to Resilience.

[209] Luís Sérgio, Ozório Valentim, Vital De Oliveira, Ribeiro Filho, Cristiane Maria, and Tranquillini Rezende. Climate action plan of the state de Sao Paulo. *Sao Paulo Epidemiological Bulletin*, 19(217):1–6, 2022.

[210] SIMA. Diretrizes e Ações Estratégicas. Plano de Ação Climática do Estado de São Paulo Net Zero 2050. Versão Preliminar para Consulta Pública. Julho de 2021. 2021.

[211] São Paulo Government. São Paulo State Climate Action plan- Net Zero 2050: Guidelines and Strategic Actions. Technical report, 2021.

[212] United Nations. Refresca São Paulo - Adapting Slums to Heat Islands Stresses and Water Insecurity - Brazil — UNFCCC.

[213] The City of São Paulo. City of São Paulo Strategic Master Plan. pages 1–24, 2014.

[214] CGE - Centro de Gerenciamento de Emergências Climáticas — São Paulo.

[215] Michelle L. Bell, Marie S. O'Neill, Nalini Ranjit, Victor H. Borja-Aburto, Luis A. Cifuentes, and Nelson C. Gouveia. Vulnerability to heat-related mortality in Latin America: A case-crossover study in São Paulo, Brazil, Santiago, Chile and Mexico City, Mexico. *International Journal of Epidemiology*, 37(4):796–804, 8 2008.

[216] Aina Roca-Barceló, Daniela Fecht, Monica Pirani, Frédéric B. Piel, Adelaide C. Nardocci, and Paolo Vineis. Trends in Temperature-associated Mortality in São Paulo (Brazil) between 2000 and 2018: an Example of Disparities in Adaptation to Cold and Heat. *Journal of Urban Health*, 99(6):1012–1026, 12 2022.

[217] Veronika Huber, Cristina Peña Ortiz, David Gallego Puyol, Stefan Lange, and Francesco Sera. Evidence of rapid adaptation integrated into projections of temperature-related excess mortality. *Environmental Research Letters*, 17(4), 2022.

[218] Brazilian Public Health System. São Paulo - DATASUS.

[219] Instituto de Astronomia Geofísica e Ciências Atmosféricas - Universidade de São Paulo. Estação Meteorológica - IAG/USP.

[220] Geofísica e Ciências Astmosféricas (IAG) da Universidade de São Paulo (USP) Estação Meteorológica (EM) do Instituto de Astronomia. RELATÓRIO TÉCNICO DA ESTAÇÃO METEOROLÓGICA DO IAG / USP Nº1 : Instrumentos e Procedimentos. Technical report, 2010.

[221] Stéphanie Bergmann, Bixia Li, Eva Pilot, Renchao Chen, Boguang Wang, and Jun Yang. Effect modification of the short-term effects of air pollution on morbidity by season: A systematic review and meta-analysis. *Science of the Total Environment*, 716:136985, 2020.

[222] Christopher J Gordon. Role of environmental stress in the physiological response to chemical toxicants. *Environmental Research*, 92:1–7, 2003.

[223] Anna Makri and Nikolaos I Stilianakis. Vulnerability to air pollution health effects. *Int. J. Hyg. Environ.-Health*, 211:326–336, 2008.

[224] CETESB. CETESB - Companhia Ambiental do Estado de São Paulo.

[225] Assembleia Legislativa do Estado de São Paulo. Lei nº 16.337, de 14 de dezembro de 2016.

[226] Assambleia Legislativa do Estado de Sao Paulo. Decreto n.59.113, de 23 de Abril de 2013 - Estabelece novos padrões de qualidade do ar e dá providências correlatas.

[227] CETESB - Companhia Ambiental do Estado de São Paulo. Classificacao expedita da representatividade espacial das estacoes de monitoramento da qualidade do ar da cetesb no estado de Sao Paolo. Technical report.

[228] Weiwei Yu, Kerrie Mengersen, Xiaoyu Wang, Xiaofang Ye, Yuming Guo, Xiaochuan Pan, and Shilu Tong. Daily average temperature and mortality among the elderly: a meta-analysis and systematic review of epidemiological evidence. *International journal of biometeorology*, 56(4):569–581, 7 2012.

[229] Rebecca Lindsey. Climate Variability: Oceanic Niño Index — NOAA Climate.gov, 2009.

[230] Yunquan Zhang, Yong Yu, Minjin Peng, Runtang Meng, Kejia Hu, and Chuanhua Yu. Temporal and seasonal variations of mortality burden associated with hourly temperature variability: A nationwide investigation in England and Wales. *Environment International*, 115:325–333, 6 2018.

[231] Hirotogu Akaike. Information Theory and an Extension of the Maximum Likelihood Principle. In Emanuel Parzen, Kunio Tanabe, and Genshiro Kitagawa, editors, *Selected Papers of Hirotugu Akaike*, pages 199–213. Springer New York, New York, NY, 1998.

[232] Organización Mundial de la Salud (OMS). WHO global air quality guidelines. *Particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide*, pages 1–360, 2021.

[233] Julia R. Barrett. News — Science Selections Increased Minimum Mortality Temperature in France. *Science Selections*, 123(7):184, 2015.

[234] F. Follos, C. Linares, J. M. Vellón, J. A. López-Bueno, M. Y. Luna, G. Sánchez-Martínez, and J. Díaz. The evolution of minimum mortality temperatures as an indicator of heat adaptation: The cases of Madrid and Seville (Spain). *Science of the Total Environment*, 747:141259, 2020.

[235] Daniel Oudin Åström, Andreas Tornevi, Kristie L. Ebi, Joacim Rocklöv, and Bertil Forsberg. Evolution of Minimum Mortality Temperature in Stockholm, Sweden, 1901–2009. *Environmental Health Perspectives*, 124(6):740–744, 2016.

[236] Adriano Massuda, Thomas Hone, Fernando Antonio Gomes Leles, Marcia C De Castro, and Rifat Atun. The Brazilian health system at crossroads: progress, crisis and resilience. *BMJ Glob Health*, 3:829, 2018.

[237] Deborah Carvalho Malta, Luisa Sorio Flor, Ísis Eloah Machado, Mariana Santos Felisbino-Mendes, Luisa Campos Caldeira Brant, Antonio Luiz Pinho Ribeiro, Renato Azeredo Teixeira, Eduardo Marques Macário, Marissa B. Reitsma, Scott Glenn, Mohsen Naghavi, and Emmanuela Gakidou. Trends in prevalence and mortality burden attributable to smoking, Brazil and federated units, 1990 and 2017. *Population Health Metrics 2020 18:1*, 18(1):1–15, 9 2020.

[238] Thomas Hone, Andrew J. Mirelman, Davide Rasella, Rômulo Paes-Sousa, Mauricio L. Barreto, Rudi Rocha, and Christopher Millett. Effect of economic recession and impact of health and social protection expenditures on adult mortality: a longitudinal analysis of 5565 Brazilian municipalities. *The Lancet Global Health*, 7(11):e1575–e1583, 2019.

[239] Arlindo Kamimura. Prospective of Brazilian Household Air Conditioning Energy Consumption and Related Carbon Emission : A Non Linear Numerical Approach Prospective of Brazilian Household Air Conditioning Energy Consumption and Related Carbon Emission : A Non Linear Numerica. (January):1–9, 2019.

[240] Jalonne L White-Newsome, Brisa N Sánchez, Edith A Parker, J Timothy Dvonch, Zhenzhen

Zhang, and Marie S. O'Neill. Assessing heat-adaptive behaviors among older, urban-dwelling adults. *Maturitas*, 70(1):85–91, 2011.

[241] Maria Angela Vigotti, Vito M.R. Muggeo, and Rosanna Cusimano. The effect of birthplace on heat tolerance and mortality in Milan, Italy, 1980-1989. *International Journal of Biometeorology*, 50(6):335–341, 7 2006.

[242] Caio Sérgio Pereira de Araújo, Ivis Andrei Campos e. Silva, Matteo Ippolito, and Ceres Duarte Guedes Cabral de Almeida. Evaluation of air temperature estimated by ERA5-Land reanalysis using surface data in Pernambuco, Brazil. *Environmental Monitoring and Assessment*, 194(5), 2022.

[243] Governo do Estado do Sao Paulo. Diário Oficial da Cidade de São Paulo - Edição de sexta-feira, 13 de maio de 2016 .

[244] Èrica Martínez-Solanas and Xavier Basagaña. Temporal changes in temperature-related mortality in Spain and effect of the implementation of a Heat Health Prevention Plan. 2018.

[245] Garth Heutel, Nolan H. Miller, and David Molitor. Adaptation and the Mortality Effects of Temperature across U.S. Climate Regions. *The Review of Economics and Statistics*, 103(4):740–753, 10 2021.

[246] Dariya Ordanovich, Aurelio Tobías, and Diego Ramiro. Temporal variation of the temperature-mortality association in Spain: a nationwide analysis. *Environmental health : a global access science source*, 22(1), 12 2023.

[247] Dimitris Evangelopoulos, Antonis Analitis, Christos Giannakopoulos, and Klea Katsouyanni. Does climatic zone of birth modify the temperature-mortality association of London inhabitants during the warm season? A time-series analysis for 2004–2013. *Environmental Research*, 193:110357, 2 2021.

[248] K. Katsouyanni, A. Pantazopoulou, G. Touloumi, L. Tselepidaki, K. Moustris, D. Asimakopoulos, G. Poulopoulou, and D. Trichopoulos. Evidence for interaction between air pollution and high temperature in the causation of excess mortality. *Archives of Environmental Health*, 48(4):235–242, 1993.

[249] Jessie P. Buckley, Jonathan M. Samet, and David B. Richardson. Does air pollution confound studies of temperature? *Epidemiology*, 25(2):242–245, 3 2014.

[250] Jian Peng, Yaxin Hu, Jianquan Dong, Qianyuan Liu, and Yanxu Liu. Quantifying spatial morphology and connectivity of urban heat islands in a megacity: A radius approach. *Science of the Total Environment*, 714:136792, 2020.

[251] Masoud Ghandehari, Thorsten Emig, and Milad Aghamohamadnia. Surface temperatures in New York City: Geospatial data enables the accurate prediction of radiative heat transfer. *Scientific Reports*, 8(1):1–10, 2018.

[252] Ioannis Kousis, Ilaria Pigliautile, and Anna Laura Pisello. Intra-urban microclimate investigation in urban heat island through a novel mobile monitoring system. *Scientific Reports*, 11(1):1–17, 2021.

[253] Kejia Hu, Yuming Guo, Stefan Hochrainer-stigler, Wei Liu, Linda See, Xuchao Yang, Jieming Zhong, Fangrong Fei, Feng Chen, Yunquan Zhang, Qi Zhao, Gongbo Chen, Qian Chen, Yizhe Zhang, Tingting Ye, Lu Ma, Shanshan Li, and Jiaguo Qi. Evidence for Urban – Rural Disparity in Temperature – Mortality Relationships in Province, Zhejiang. *Environmental Health Perspectives*, 127(March):037001, 2019.

[254] Leah H. Schinasi, Tarik Benmarhnia, and Anneclaire J. De Roos. Modification of the association between high ambient temperature and health by urban microclimate indicators: A systematic review and meta-analysis. *Environmental Research*, 161(September 2017):168–180, 2018.

[255] NASA. Earthdata — Earthdata.

[256] Tom G. Farr, Paul A. Rosen, Edward Caro, Robert Crippen, Riley Duren, Scott Hensley, Michael Kobrick, Mimi Paller, Ernesto Rodriguez, Ladislav Roth, and David Seal. The shuttle radar topography mission. *Rev. Geophys.*, RG2004(2005):65–77, 2007.

[257] Ying Qu, Shunlin Liang, Qiang Liu, Tao He, Suhong Liu, and Xiaowen Li. Mapping surface broadband albedo from satellite observations: A review of literatures on algorithms and products. *Remote Sensing*, 7(1):990–1020, 2015.

[258] USGS. LP DAAC - MCD43A3 v061.

[259] A$\rho\rho$EEARS.

[260] Jae Dong Jang, A. A. Viau, and F. Anctil. Neural network estimation of air temperatures from AVHRR data. *International Journal of Remote Sensing*, 25(21):4541–4554, 2004.

[261] Christelle Vancutsem, Pietro Ceccato, Tufa Dinku, and Stephen J. Connor. Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa. *Remote Sensing of Environment*, 114(2):449–465, 2010.

[262] NASA. LP DAAC - MOD13A2_v006.

[263] Zhengming Wan. *Collection-6 MODIS Land Surface Temperature Products Users' Guide*. Santa Barbara, 2013.

[264] Pierre C. Guillevic, James C. Biard, Glynn C. Hulley, Jeffrey L. Privette, Simon J. Hook, Albert Olioso, Frank M. Göttsche, Robert Radocinski, Miguel O. Román, Yunyue Yu, and Ivan Csiszar. Validation of Land Surface Temperature products derived from the Visible Infrared Imaging Radiometer Suite (VIIRS) using ground-based and heritage satellite measurements. *Remote Sensing of Environment*, 154:19–37, 2014.

[265] D. J. Ghent, G. K. Corlett, F. M. Göttsche, and J. J. Remedios. Global Land Surface Temperature From the Along-Track Scanning Radiometers. *Journal of Geophysical Research: Atmospheres*, 122(22):167–12, 2017.

[266] Jaime Aguilar-Lome, Raúl Espinoza-Villar, Jhan Carlo Espinoza, Joel Rojas-Acuña, Bram Leo Willems, and Walter Martín Leyva-Molina. Elevation-dependent warming of land surface temperatures in the Andes assessed using MODIS LST time series (2000–2017). *International Journal of Applied Earth Observation and Geoinformation*, 77(January):119–128, 2019.

[267] Shushi Peng, Shilong Piao, Philippe Ciais, Pierre Friedlingstein, Catherine Ottle, François Marie Bréon, Huijuan Nan, Liming Zhou, and Ranga B. Myneni. Response to comment on "Surface urban heat island across 419 global big cities". *Environmental Science and Technology*, 46(12):6889–6890, 2012.

[268] Bin Zhou, Evyatar Erell, Ian Hough, Jonathan Rosenblatt, Allan C. Just, Victor Novack, and Itai Kloog. Estimating near-surface air temperature across Israel using a machine learning based hybrid approach. *International Journal of Climatology*, 40(14):6106–6121, 2020.

[269] Shilo Shiff, David Helman, and Itamar M. Lensky. Worldwide continuous gap-filled MODIS land surface temperature dataset. *Scientific Data*, 8(1):1–10, 2021.

[270] Zhaowu Yu, Gaoyuan Yang, Shudi Zuo, Gertrud Jørgensen, Motoya Koga, and Henrik Vejre. Critical review on the cooling effect of urban blue-green space: A threshold-size perspective. *Urban Forestry & Urban Greening*, 49:126630, 3 2020.

[271] Comitê da Bacia Hidrográfica do Alto Tietê (CBH-AT). Datasets.

[272] Natural Earth - Free vector and raster map data at 1:10m, 1:50m, and 1:110m scales.

[273] Open Spatial Demographic Data and Research - WorldPop.

[274] Forrest R. Stevens, Andrea E. Gaughan, Catherine Linard, and Andrew J. Tatem. Disaggregating census data for population mapping using Random forests with remotely-sensed and ancillary data. *PLoS ONE*, 10(2):1–22, 2015.

[275] Xiao Zhang, Liangyun Liu, Changshan Wu, Xidong Chen, Yuan Gao, Shuai Xie, and Bing Zhang. Development of a global 30m impervious surface map using multisource and multitemporal remote sensing datasets with the Google Earth Engine platform. *Earth System Science Data*, 12(3):1625–1648, 2020.

[276] ESA. Land Cover ESA Climate Change Initiative project.

[277] GEOSAMPA: Sistema de Consulta do Mapa Digital da Cidade de São Paulo :.

[278] Earth Engine Data Catalog. USGS Landsat 7 Level 2, Collection 2, Tier 1 — Earth Engine Data Catalog — Google Developers, 2022.

[279] J.B. James MacQueen et al. Some methods for classification and analysis of multivariate observations. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, 1(14):281–297, 1 1967.

[280] Google Earth Engine.

[281] Florian Gerber, Reinhard Furrer, Gabriela Schaepman-Strub, Rogier de Jong, and Michael E. Schaepman. Predicting missing values in spatio-temporal satellite data. *IEEE Transactions on Geoscience and Remote Sensing*, 56:1–36, 2016.

[282] Suranjana Saha, Shrinivas Moorthi, Xingren Wu, Jiande Wang, Sudhir Nadiga, Patrick Tripp, David Behringer, Yu Tai Hou, Hui Ya Chuang, Mark Iredell, Michael Ek, Jesse Meng, Rongqian Yang, Malaquías Peña Mendez, Huug Van Den Dool, Qin Zhang, Wanqiu Wang, Mingyue Chen, and Emily Becker. Climate Forecast System Version 2 (CFSv2) Operational Forecasts. *Journal of Climate*, 27(6):2185–2208, 2014.

[283] Tao Zhang, Yuyu Zhou, Zhengyuan Zhu, Xiaoma Li, and Ghassem R. Asrar. A global seamless 1km resolution daily land surface temperature dataset (2003-2020). *Earth System Science Data*, 14(2):651–664, 2022.

[284] Florian Gerber. Package 'gapfill' Type Package Title Fill Missing Values in Satellite Data Depends R (¿= 3.1), ggplot2 (¿= 2.2.1), 2022.

[285] Leo Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996.

[286] Hanna Meyer, Christoph Reudenbach, Tomislav Hengl, Marwan Katurji, and Thomas Nauss. Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. *Environmental Modelling and Software*, 101:1–9, 2018.

[287] Itai Kloog, Francesco Nordio, Brent A. Coull, and Joel Schwartz. Predicting spatiotemporal mean air temperature using MODIS satellite surface temperature measurements across the Northeastern USA. *Remote Sensing of Environment*, 150:132–139, 2014.

[288] J. A. Voogt and T. R. Oke. Effects of urban surface geometry on remotely-sensed surface temperature. *International Journal of Remote Sensing*, 19(5):895–920, 1998.

[289] Xu Lian, Zhenzhong Zeng, Yitong Yao, Shushi Peng, Kaicun Wang, and Shilong Piao. Spatiotemporal variations in the difference between satellite-observed daily maximum land surface temperature and station-based daily maximum near-surface air temperature. *Journal of Geophysical Research*, 122(4):2254–2268, 2017.

[290] Hadas Saaroni, Eyal Ben-Dor, Arieh Bitan, and Oded Potchter. Spatial distribution and microscale characteristics of the urban heat island in Tel-Aviv, Israel. *Landscape and Urban Planning*, 48(1-2):1–18, 2000.

[291] Gang Fu, Zhenxi Shen, Xianzhou Zhang, Peili Shi, Yangjian Zhang, and Jianshuang Wu. Estimating air temperature of an alpine meadow on the Northern Tibetan Plateau using MODIS land surface temperature. *Acta Ecologica Sinica*, 31(1):8–13, 2011.

[292] Cheolhee Yoo, Jungho Im, Seonyoung Park, and Lindi J. Quackenbush. Estimation of daily maximum and minimum air temperatures in urban landscapes using MODIS time series satellite data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 137:149–162, 2018.

[293] Augusto Cezar Lima Do Nascimento, Emerson Galvani, João Paulo Assis Gobo, and Cássio Arthur Wollmann. Comparison between Air Temperature and Land Surface Temperature for the City of São Paulo, Brazil. *Atmosphere*, 13(3):1–21, 2022.

[294] Hung Chak Ho, Anders Knudby, Paul Sirovyak, Yongming Xu, Matus Hodul, and Sarah B. Henderson. Mapping maximum urban air temperature on hot summer days. *Remote Sensing of Environment*, 154:38–45, 2014.

[295] Yuan Z. Yang, Wen H. Cai, and Jian Yang. Evaluation of MODIS land surface temperature data to estimate near-surface air temperature in Northeast China. *Remote Sensing*, 9(5):1–19, 2017.

[296] Wenbin Zhu, Aifeng Lu, and Shaofeng Jia. Estimation of daily maximum and minimum air temperature using MODIS land surface temperature products. *Remote Sensing of Environment*, 130:62–73, 2013.

[297] Benedicte Dousset. AVHRR-derived cloudiness and surface temperature patterns over the Los Angeles area and their relationship to land use. In *Proceedings of IGARSS-89*, pages 2132–2137, New York, NY, 1989. IEEE.

[298] Jin Zou, Ning Lu, Hou Jiang, Jun Qin, Ling Yao, Ying Xin, and Fenzhen Su. Performance of air temperature from ERA5-Land reanalysis in coastal urban agglomeration of Southeast China. *Science of the Total Environment*, 828:154459, 2022.

[299] Pierre Masselot, Malcolm Mistry, Jacopo Vanoli, Rochelle Schneider, Tamara Iungman, David Garcia-Leon, Juan-Carlos Ciscar, Luc Feyen, Hans Orru, Aleš Urban, Susanne Breitner, Veronika Huber, Alexandra Schneider, Evangelia Samoli, Massimo Stafoggia, Francesca de'Donato, Shilpa Rao, Ben Armstrong, Mark Nieuwenhuijsen, Ana Maria Vicedo-Cabrera, Antonio Gasparrini, Souzana Achilleos, Jan Kyselý, Ene Indermitte, Jouni J.K. Jaakkola, Niilo Ryti, Mathilde Pascal, Klea Katsouyanni, Antonis Analitis, Patrick Goodman, Ariana Zeka, Paola Michelozzi, Danny Houthuijs, Caroline Ameling, Shilpa Rao, Susana das Neves Pereira da Silva, Joana Madureira, Iulian-Horia Holobaca, Aurelio Tobias, Carmen Íñiguez, Bertil Forsberg, Christofer Åström, Martina S. Ragettli, Antonis Analitis, Klea Katsouyanni, First name Surname, Sofia Zafeiratou, Liliana Vazquez Fernandez, Ana Monteiro, Masna Rai, Siqi Zhang, and Kristin Aunan. Excess mortality attributed to heat and cold: a health impact assessment study in 854 cities in Europe. *The Lancet Planetary Health*, pages 271–281, 2023.

[300] M. A. Friedl and F. W. Davis. Sources of variation in radiometric surface temperature over a tallgrass prairie. *Remote Sensing of Environment*, 48(1):1–17, 1994.

[301] SN Goward and AS Hope. Evapotranspiration from combined reflected solar and emitted terrestrial radiation: preliminary results from AVHRR data. *Advances in Space Research*, 9(7):239–249, 1989.

[302] Xiudi Zhu, Qiang Zhang, Chong Yu Xu, Peng Sun, and Pan Hu. Reconstruction of high spatial resolution surface air temperature data across China: A new geo-intelligent multisource data-based machine learning technique. *Science of the Total Environment*, 665:300–313, 2019.

[303] Hongbo Zhang, Fan Zhang, Ming Ye, Tao Che, and Guoqing Zhang. Estimating daily air temperatures over the Tibetan Plateau by dynamically integrating MODIS LST data. *Journal of Geophysical Research Atmospheres*, 121(11):11,425, 2016.

[304] Iván Gutiérrez-Avila, Kodi B. Arfer, Sandy Wong, Johnathan Rush, Itai Kloog, and Allan C. Just. A spatiotemporal reconstruction of daily ambient temperature using satellite data in the Megalopolis of Central Mexico from 2003 to 2019. *International Journal of Climatology*, 41(8):4095–4111, 2021.

[305] Itai Kloog, Alexandra A. Chudnovsky, Allan C. Just, Francesco Nordio, Petros Koutrakis, Brent A. Coull, Alexei Lyapustin, Yujie Wang, and Joel Schwartz. A new hybrid spatio-temporal model for estimating daily multi-year PM2.5 concentrations across northeastern USA using high resolution aerosol optical depth data. *Atmospheric Environment*, 95:581–590, 10 2014.

[306] Liuhua Shi, Pengfei Liu, Itai Kloog, Mihye Lee, Anna Kosheleva, and Joel Schwartz. Estimating daily air temperature across the Southeastern United States using high-resolution satellite data: A statistical modeling study. *Environmental Research*, 146:51–58, 2016.

[307] Itai Kloog, Francesco Nordio, Johanna Lepeule, Andrea Padoan, Mihye Lee, Annick Auffray, and Joel Schwartz. Modelling spatio-temporally resolved air temperature across the complex geo-climate area of France using satellite-derived land surface temperature data. *International Journal of Climatology*, 37(1):296–304, 1 2017.

[308] Adar Rosenfeld, Michael Dorman, Joel Schwartz, Victor Novack, Allan C. Just, and Itai Kloog. Estimating daily minimum, maximum, and mean near surface air temperature using hybrid satellite models across Israel. *Environmental Research*, 159(March):297–312, 2017.

[309] Jin Li, Andrew D. Heap, Anna Potter, and James J. Daniell. Application of machine learning methods to spatial interpolation of environmental variables. *Environmental Modelling and Software*, 26(12):1647–1659, 2011.

[310] Sahar Mohsenzadeh Karimi, Ozgur Kisi, Mohammad Porrajabali, Fatemeh Rouhani-Nia, and Jalal Shiri. Evaluation of the support vector machine, random forest and geo-statistical methodologies for predicting long-term air temperature. *ISH Journal of Hydraulic Engineering*, 26(4):376–386, 2020.

[311] Yongming Xu, Anders Knudby, and Hung Chak Ho. Estimating daily maximum air temperature from MODIS in British Columbia, Canada. *International Journal of Remote Sensing*, 35(24):8108–8121, 2014.

[312] Sergio M. Vicente-Serrano, M. Angel Saz-Sánchez, and José M. Cuadrat. Comparative analysis of interpolation methods in the middle Ebro Valley (Spain): Application to annual precipitation and temperature. *Climate Research*, 24(2):161–180, 2003.

[313] E. N. Florio, S. R. Lele, Y. Chi Chang, R. Sterner, and G. E. Glass. Integrating AVHRR satellite data and NOAA ground observations to predict surface air temperature: A statistical approach. *International Journal of Remote Sensing*, 25(15):2979–2994, 2004.

[314] Fabio Veronesi and Calogero Schillaci. Comparison between geostatistical and machine learning

models as predictors of topsoil organic carbon with a focus on local uncertainty estimation. *Ecological Indicators*, 101(February):1032–1044, 2019.

[315] Hirofumi Hashimoto, Weile Wang, Forrest S. Melton, Adam L. Moreno, Sangram Ganguly, Andrew R. Michaelis, and Ramakrishna R. Nemani. High-resolution mapping of daily climate variables by aggregating multiple spatial data sets with the random forest algorithm over the conterminous United States. *International Journal of Climatology*, 39(6):2964–2983, 2019.

[316] Tomislav Hengl, Madlene Nussbaum, Marvin N. Wright, Gerard B.M. Heuvelink, and Benedikt Gräler. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, 2018(8), 2018.

[317] Ran Pelta and Alexandra A. Chudnovsky. Spatiotemporal estimation of air temperature patterns at the street level using high resolution satellite imagery. *Science of the Total Environment*, 579:675–684, 2017.

[318] Chris Funk, Pete Peterson, Seth Peterson, Shraddhanand Shukla, Frank Davenport, Joel Michaelsen, Kenneth R. Knapp, Martin Landsfeld, Gregory Husak, Laura Harrison, James Rowland, Michael Budde, Alex Meiburg, Tufa Dinku, Diego Pedreros, and Nicholas Mata. A high-resolution 1983–2016 TMAX climate data record based on infrared temperatures and stations by the climate hazard center. *Journal of Climate*, 32(17):5639–5658, 2019.

[319] Gerald Mills. Progress toward sustainable settlements: A role for urban climatology. *Theoretical and Applied Climatology*, 84(1-3):69–76, 2006.

[320] Lei Wang, Toshio Koike, Kun Yang, and Pat Jen Feng Yeh. Assessment of a distributed biosphere hydrological model against streamflow and MODIS land surface temperature in the upper Tone River Basin. *Journal of Hydrology*, 377(1-2):21–34, 10 2009.

[321] Tini Garske, Neil M. Ferguson, and Azra C. Ghani. Estimating Air Temperature and Its Influence on Malaria Transmission across Africa. *PLOS ONE*, 8(2):e56487, 2 2013.

[322] World Health Organization (WHO). Determinants of health, 2019.

[323] Ministério do Trabalho e Emprego Governo Federal do Brasil. *Classificação Brasileira de Ocupaçoes*, volume 1. 2010.

[324] SMDUR - Secretaria Municipal de Desenvolvimento Urbano.

[325] Atlas Brasil. Atlas do Desenvolvimento Humano no Brasil.

[326] IPEA; UNDP; Fundacao Joao Pinheiro. *Human Development Atlas in the Brazilian Metropolitan Regions*. 2014.

[327] IPEA; UNDP; Fundacao Jao Pinheiro. NOTA TÉCNICA No. 01/2022 Aspectos Metodológicos para o cálculo do Índice de Desenvolvimento Humano Municipal. Technical Report 01, 2022.

[328] UNDP — IPEA — FJP. Human Development Atlas in the Brazilian Metropolitan Regions. Technical report, 2014.

[329] Terry M Therneau, Thomas Lumley, Atkinson Elizabeth, and Crowson Cynthia. Package ' survival ', 2023.

[330] Peter Mehnert, Peter Bröde, and Barbara Griefahn. Gender-related difference in sweat loss and its impact on exposure limits to heat stress. *International Journal of Industrial Ergonomics*, 29(6):343–351, 2002.

[331] D. F. Archer, D. W. Sturdee, R. Baber, T. J. De Villiers, A. Pines, R. R. Freedman, A. Gompel, M. Hickey, M. S. Hunter, R. A. Lobo, M. A. Lumsden, A. H. MacLennan, P. Maki, S. Palacios, D. Shah, P. Villaseca, and M. Warren. Menopausal hot flushes and night sweats: Where are we now? *Climacteric*, 14(5):515–528, 2011.

[332] Yvette van Steen, Anna Maria Ntarladima, Rick Grobbee, Derek Karssenberg, and Ilonca Vaartjes. Sex differences in mortality after heat waves: are elderly women at higher risk?, 1 2019.

[333] Joan Ballester, Marcos Quijal-zamorano, Raúl Fernando, Méndez Turrubiates, Ferran Pegenaute, François R Herrmann, Jean Marie Robine, and Xavier Basagaña. Heat-related mortality in Europe during the summer of 2022. 2023.

[334] Yiqun Ma, Lian Zhou, and Kai Chen. Burden of cause-specific mortality attributable to heat and cold: A multicity time-series study in Jiangsu Province, China. *Environment International*, 144(172):105994, 2020.

[335] Noah Scovronick and Ben Armstrong. The impact of housing type on temperature-related mortality in South Africa, 1996-2015. *Environmental Research*, 113(2012):46–51, 2012.

[336] Thaddaeus Egondi, Catherine Kyobutungi, Sari Kovats, Kanyiva Muindi, Remare Ettarh, and Joacim Rocklöv. Time-series analysis of weather and mortality patterns in Nairobi's informal settlements. *Global Health Action*, 5(1):19065, 2012.

[337] Vijendra Ingole, Sari Kovats, Barbara Schumann, Shakoor Hajat, Joacim Rocklov, Sanjay Juvekar, and Ben Armstrong. Socioenvironmental factors associated with heat and cold-related mortality in Vadu HDSS, western India: a population-based case-crossover study. *International journal of biometeorology*, 61(10):1797–1804, 10 2017.

[338] Edelci Nunes da Silva and Helena Ribeiro. Alterações da temperatura em ambientes externos de favela e desconforto térmico. *Revista de Saúde Pública*, 40(4):663–670, 2006.

[339] Helen L. Macintyre, Clare Heaviside, Xiaoming Cai, and Revati Phalkey. The winter urban heat island: Impacts on cold-related mortality in a highly urbanized European region for present and future climate. *Environment International*, 154(January):106530, 2021.

[340] Emmanuel A Odame, Ying Li, Shimin Zheng, Ambarish Vaidyanathan, and Ken Silver. Assessing Heat-Related Mortality Risks among Rural Populations: A Systematic Review and Meta-Analysis of Epidemiological Evidence. *International journal of environmental research and public health*, 15(8), 7 2018.

[341] Anthony J McMichael, Paul Wilkinson, R Sari Kovats, Sam Pattenden, Shakoor Hajat, Ben Armstrong, Nitaya Vajanapoom, Emilia M Niciu, Hassan Mahomed, Chamnong Kingkeow, Mitja Kosnik, Marie S O'Neill, Isabelle Romieu, Matiana Ramirez-Aguilar, Mauricio L Barreto, Nelson Gouveia, and Bojidar Nikiforov. International study of temperature, heat and urban mortality: the 'ISOTHURM' project. *International journal of epidemiology*, 37(5):1121–1131, 10 2008.

[342] Luis Bértola and Jeffrey Williamson. Has Latin American inequality changed direction?: Looking over the long run. *Has Latin American Inequality Changed Direction?: Looking Over the Long Run*, pages 1–419, 2017.

[343] Ben G. Armstrong. Fixed factors that modify the effects of time-varying factors: Applying the case-only approach. *Epidemiology*, 14(4):467–472, 2003.

[344] Mika R. Moran, Usama Bilal, Iryna Dronova, Yang Ju, Nelson Gouveia, Waleska Teixeira Caiaffa, Amélia Augusta de Lima Friche, Kari Moore, J. Jaime Miranda, and Daniel A. Rodríguez. The equigenic effect of greenness on the association between education with life expectancy and mortality in 28 large Latin American cities. *Health and Place*, 72(June), 2021.

[345] Sida Liu, Emily Yang Ying Chan, William Bernard Goggins, and Zhe Huang. The mortality risk and socioeconomic vulnerability associated with high and low temperature in Hong Kong. *International Journal of Environmental Research and Public Health*, 17(19):1–14, 2020.

[346] Rongbin Xu, Qi Zhao, Micheline S.Z.S. Coelho, Paulo H.N. Saldiva, Michael J Abramson, Shanshan Li, and Yuming Guo. Socioeconomic inequality in vulnerability to all-cause and cause-specific hospitalisation associated with temperature variability: a time-series study in 1814 Brazilian cities. *The Lancet Planetary Health*, 4(12):e566–e576, 2020.

[347] Eleni Oikonomou, Michael Davies, Anna Mavrogianni, Phillip Biddulph, Paul Wilkinson, and Maria Kolokotroni. Modelling the relative importance of the urban heat island and the thermal quality of dwellings for overheating in London. *Building and Environment*, 57:223–238, 11 2012.

[348] Francesco Sera, Masahiro Hashizume, Yasushi Honda, Eric Lavigne, Joel Schwartz, Antonella Zanobetti, Aurelio Tobias, Carmen Iñiguez, Ana M. Vicedo-Cabrera, Marta Blangiardo, Ben Armstrong, and Antonio Gasparrini. Air conditioning and heat-related mortality: A multi-country longitudinal study. *Epidemiology*, 31(6):779–787, 2020.

[349] Liuhua Shi, Pengfei Liu, Yan Wang, Antonella Zanobetti, Anna Kosheleva, Petros Koutrakis, and Joel Schwartz. Chronic effects of temperature on mortality in the Southeastern USA using satellite-based exposure metrics. *Scientific Reports*, 6(November 2015):1–8, 2016.

[350] I M Heid, H Küchenhoff, J Miles, L Kreienbrock, and H E Wichmann. Two dimensions of measurement error: classical and Berkson error in residential radon exposure assessment. *Journal of exposure analysis and environmental epidemiology*, 14(5):365–377, 9 2004.

[351] Janet Ige-Elegbede, Paul Pilkington, Judy Orme, Ben Williams, Emily Prestwood, Daniel Black, and Laurence Carmichael. Designing healthier neighbourhoods: a systematic review of the impact of the neighbourhood design on health and wellbeing. *Cities and Health*, 6(5):1004–1019, 2022.

[352] Jeffrey Raven, Brian Stone, Gerald Mills, Joel Towers, Lutz Katzschner, Mattia Federico Leone, Pascaline Gaborit, Matei Georgescu, Maryam Hariri, James Lee, Jeffrey LeJava, Ayyoob Sharifi, Cristina Visconti, and Andrew Rudd. Urban Planning and Urban Design. In *Climate Change and Cities: Second Assessment Report of the Urban Climate Change Research Network*, pages 139–172. Cambridge University Press, 1 2018.

[353] Audrey de Nazelle, Charlotte J. Roscoe, Aina Roca-Barcel, Giselle Sebag, Gudrun Weinmayr, Carlos Dora, Kristie L. Ebi, Mark J. Nieuwenhuijsen, and Maya Negev. Urban climate policy and action through a health lens—an untapped opportunity. *International Journal of Environmental Research and Public Health*, 18(23):1–9, 2021.

[354] EPA. Smart Growth — US EPA.

[355] Andrew MacLachlan, Eloise Biggs, Gareth Roberts, and Bryan Boruff. Sustainable City Planning: A Data-Driven Approach for Mitigating Urban Heat. *Frontiers in Built Environment*, 6(January):1–12, 2021.

[356] Zander S. Venter, Oscar Brousse, Igor Esau, and Fred Meier. Hyperlocal mapping of urban air

temperature using remote sensing and crowdsourced weather data. *Remote Sensing of Environment*, 242(January):111791, 2020.

[357] Hanna Meyer and Edzer Pebesma. Predicting into unknown space? Estimating the area of applicability of spatial prediction models. *Methods in Ecology and Evolution*, 12(9):1620–1633, 2021.

[358] Mathieu Marmion, Miia Parviainen, Miska Luoto, Risto K. Heikkinen, and Wilfried Thuiller. Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions*, 15(1):59–69, 2009.

[359] Mohsen B. Mesgaran, Roger D. Cousens, and Bruce L. Webber. Here be dragons: A tool for quantifying novelty due to covariate range and correlation change when projecting species distribution models. *Diversity and Distributions*, 20(10):1147–1159, 2014.

[360] Owen L. Petchey, Mikael Pontarp, Thomas M. Massie, Sonia Kéfi, Arpat Ozgul, Maja Weilenmann, Gian Marco Palamara, Florian Altermatt, Blake Matthews, Jonathan M. Levine, Dylan Z. Childs, Brian J. Mcgill, Michael E. Schaepman, Bernhard Schmid, Piet Spaak, Andrew P. Beckerman, Frank Pennekamp, and Ian S. Pearse. The ecological forecast horizon, and examples of its uses and determinants. *Ecology Letters*, 18(7):597–611, 2015.

[361] Seoncheol Park, Junhyeon Kwon, Joonpyo Kim, and Hee Seok Oh. Prediction of extremal precipitation by quantile regression forests: from SNU Multiscale Team. *Extremes*, 21(3):463–476, 2018.

[362] Nicola Gnecco, Edossa Merga Terefe, and Sebastian Engelke. Extremal Random Forests. 2023.

[363] John Gulliver, Kees De Hoogh, Anna Hansell, and Danielle Vienneau. Development and backextrapolation of NO2 land use regression models for historic exposure assessment in Great Britain. *Environmental Science and Technology*, 47(14):7804–7811, 7 2013.

[364] Katherine L. Yates, Phil J. Bouchet, M. Julian Caley, Kerrie Mengersen, Christophe F. Randin, Stephen Parnell, Alan H. Fielding, Andrew J. Bamford, Stephen Ban, A. Márcia Barbosa, Carsten F. Dormann, Jane Elith, Clare B. Embling, Gary N. Ervin, Rebecca Fisher, Susan Gould, Roland F. Graf, Edward J. Gregr, Patrick N. Halpin, Risto K. Heikkinen, Stefan Heinänen, Alice R. Jones, Periyadan K. Krishnakumar, Valentina Lauria, Hector Lozano-Montes, Laura Mannocci, Camille Mellin, Mohsen B. Mesgaran, Elena Moreno-Amat, Sophie Mormede, Emilie Novaczek, Steffen Oppel, Guillermo Ortuño Crespo, A. Townsend Peterson, Giovanni Rapacciuolo, Jason J. Roberts, Rebecca E. Ross, Kylie L. Scales, David Schoeman, Paul Snelgrove, Göran Sundblad, Wilfried Thuiller, Leigh G. Torres, Heroen Verbruggen, Lifei

Wang, Seth Wenger, Mark J. Whittingham, Yuri Zharikov, Damaris Zurell, and Ana M.M. Sequeira. Outstanding Challenges in the Transferability of Ecological Models. *Trends in Ecology and Evolution*, 33(10):790–802, 2018.

[365] Jeremy S. Hoffman, Vivek Shandas, and Nicholas Pendleton. The Effects of Historical Housing Policies on Resident Urban Areas. *Climate*, 8(12):1–15, 2020.

[366] ORORATECH: Thermal Intelligence - We Scan the World's Temperature.

[367] Destination Earth — Shaping Europe's digital future.

[368] Rochelle Schneider dos Santos, Ana M Vicedo-Cabrera, Francesco Sera, Pierre Masselot, Massimo Stafoggia, Kees de Hoogh, Itai Kloog, Stefan Reis, Massimo Vieno, and Antonio Gasparrini. A satellite-based spatio-temporal machine learning model to reconstruct daily PM2.5 concentrations across Great Britain.

[369] Carles Milà, Joan Ballester, Xavier Basagaña, Mark J. Nieuwenhuijsen, and Cathryn Tonne. Estimating daily air temperature and pollution in Catalonia: A comprehensive spatiotemporal modelling of multiple exposures. *Environmental Pollution*, 337:122501, 11 2023.

[370] Klea Katsouyanni and Dimitris Evangelopoulos. Invited Perspective: Impact of Exposure Measurement Error on Effect Estimates—An Important and Neglected Problem in Air Pollution Epidemiology. *Environmental Health Perspectives*, 130(7), 7 2022.

[371] Amanda Y Kong and Xingyou Zhang. The Use of Small Area Estimates in Place-Based Health Research. *Public Health*, 110:829–832, 2020.

[372] M. Medina-Ramón and J. Schwartz. Temperature, temperature extremes, and mortality: A study of acclimatisation and effect modification in 50 US cities. *Occupational and Environmental Medicine*, 64(12):827–833, 2007.

[373] Robbie M. Parks, James E. Bennett, Helen Tamura-Wicks, Vasilis Kontis, Ralf Toumi, Goodarz Danaei, and Majid Ezzati. Anomalously warm temperatures are associated with increased injury deaths. *Nature Medicine*, 26(1):65–70, 2020.

[374] I. Gómez-Acebo, J. Llorca, and T. Dierssen. Cold-related mortality due to cardiovascular diseases, respiratory diseases and cancer: A case-crossover study. *Public Health*, 127(3):252–258, 2012.

[375] Hannah Mason, Jemma C King, Amy E Peden, and Richard C Franklin. Systematic review of the impact of heatwaves on health service demand in Australia. *BMC Health Services Research*, 22(1):1–13, 2022.

[376] Dung Phung, Phong K. Thai, Yuming Guo, Lidia Morawska, Shannon Rutherford, and Cordia Chu. Ambient temperature and risk of cardiovascular hospitalization: An updated systematic review and meta-analysis. *Science of The Total Environment*, 550:1084–1102, 4 2016.

[377] Li Juan Huang, Jun Jing Zha, Nv Wei Cao, Hao Yue Zhou, Xiu Jie Chu, Hua Wang, Xian Bao Li, and Bao Zhu Li. Temperature might increase the hospital admission risk for rheumatoid arthritis patients in Anqing, China: a time-series study. *International Journal of Biometeorology*, 66(1):201–211, 2022.

[378] Shengzhi Sun, Francine Laden, Jaime E. Hart, Hong Qiu, Yan Wang, Chit Ming Wong, Ruby Siu Yin Lee, and Linwei Tian. Seasonal temperature variability and emergency hospital admissions for respiratory diseases: A population-based cohort study. *Thorax*, 73(10):951–958, 2018.

[379] M. Stafoggia, J. Schwartz, F. Forastiere, and C. A. Perucci. Does temperature modify the association between air pollution and mortality? A multicity case-crossover analysis in Italy. *American journal of epidemiology*, 167(12):1476–1485, 2008.

[380] Luciana Leirião, Michelle de Oliveira, Tiago Martins, and Simone Miraglia. A Multi-Pollutant and Meteorological Analysis of Cardiorespiratory Mortality among the Elderly in São Paulo, Brazil—An Artificial Neural Networks Approach. *International Journal of Environmental Research and Public Health*, 20(8), 2023.

[381] Editorial. Why a good lab website matters. *Nature Metabolism*, 5(4):533, 2023.

Appendix

6.7 Information

6.7.1 Data sharing

Project's website: Project and team websites have been shown to be a useful tool to engage with the general public [381]. Thus, I have developed a personal website using Github Pages, which encompasses various sections. The site includes four tabs: about me, the project, the data and contact us. Figure 6.1 shows the landing page for the website. The "About me" page provides information about myself, my training and professional experience. In the near future, this will also include information on my supervisors and collaborators, highlighting their expertise and involvement in the study. The "The project" tab allows users to access a concise overview of the project's rationale and gain access to each individual chapter. Upon visiting the chapter-specific pages, users will find an interactive table of contents, guiding them through the narrative. In general, the content is structured akin to a research paper, commencing with setting the context and justifying the study's importance. It subsequently describes the study's methodology and presents the results, followed by a discussion contextualizing the findings within existing literature. All information is conveyed in plain English, employing relatable examples and descriptions of the concepts discussed. The results presented mirror those published in respective scientific publications, albeit presented in a more visual and interactive format, allowing users to engage with the interface and explore the data. Under "The data" section, there are links to data sources and code. Finally, the "Contact us" tab redirects the user to a feedback form that is linked to my personal email so that I can provide immediate response and address their concerns and queries. Currently, the website has not been extensively promoted, but once it is, I will have the opportunity to measure its engagement using various metrics. These parameters will include the total number of views, the count of unique visitors, and an analysis of the most popular content on the site. Armed with this valuable data, I will be able to enhance the website's performance and pinpoint specific groups that demonstrate a keen interest in its offerings. By leveraging this information, I can effectively tailor the website to cater to the preferences and needs of the audience, ensuring a more enriching and rewarding experience for all users.

The website can be accessed here: `https://ainarb.github.io/climate_and_health/`

GitHub page: The computer code necessary for replicating the results presented in each chapter

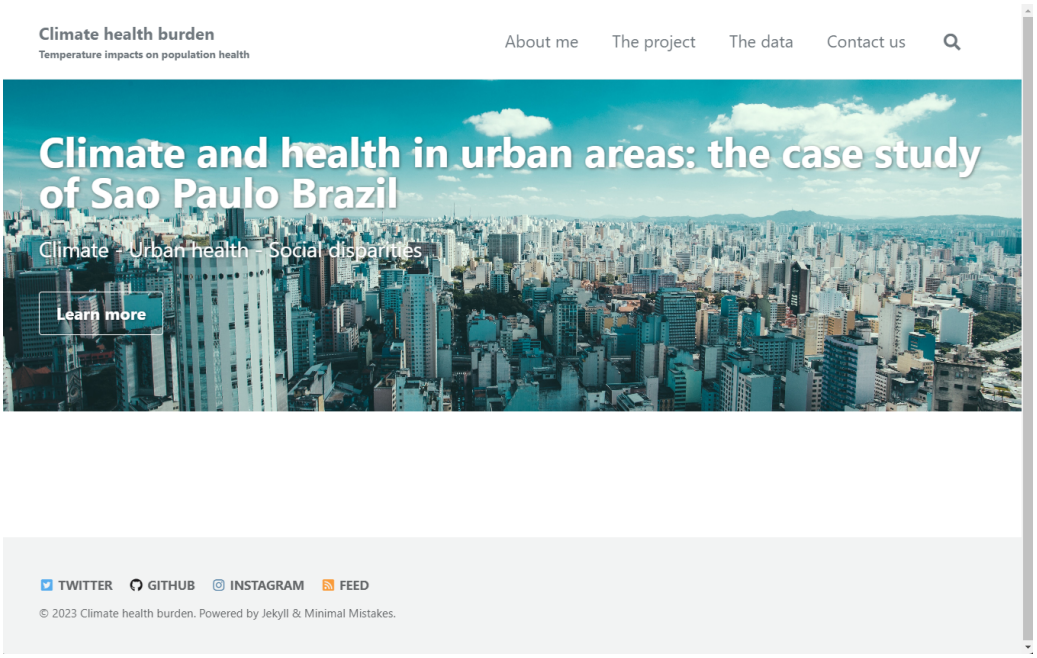


Figure 6.1: **Project's website landing page.** View of the landing page of the website for the project, with the different tabs at the top.

is available on my personal GitHub site. A dedicated repository exists for each chapter, tailored to the specific project. Although the contents may differ across repositories, there are two consistent documents: a *Readme* file and a *license* file. The *Readme* file adheres to a standardized structure, encompassing an abstract, links to the website and scientific publication relevant to the chapter, a description of the utilized data along with instructions on accessing or downloading it if not provided as part of the repository, an overview of the statistical model employed, a description of the folder structure, and a section on how to appropriately cite the associated paper. The *license* file outlines the specific type of license governing the provided information, delineating the associated copyright and sharing rights. In addition to the two files mentioned earlier, each repository includes access to the R or Python scripts utilized for the analyses. These scripts generally replicate the exact code executed to obtain the published results, with the exception of certain code lines for the data cleaning of the geo-coded mortality data, which were modified to address confidentiality concerns. Moreover, a data folder is consistently provided, containing the datasets used to generate the reported results in the paper. However, in some instances, the availability of this data is restricted due to confidentiality or data sharing restrictions from the data providers. In such cases, a synthetic dataset is made available, suitable for executing the code, along with instructions on accessing and downloading the original data, when applicable.

My GitHub repositories can be accessed here: `https://github.com/AinaRB`

6.7.2 Collaboration with the University of São Paulo

The University of São Paulo (USP), is a state-funded educational institution situated in the Brazilian state of São Paulo. Regarded as the most prominent educational establishment in the country, USP's academic pursuits encompass a wide spectrum of disciplines, including teaching, research, and university extension activities, facilitating a comprehensive array of courses across various fields of knowledge. The collaboration between Imperial College and USP began in 2016 when Imperial College received the FAPESP-Imperial College London SPRING grant to establish a partnership on big data usage and small area studies. As a Research Assistant in Epidemiology at that time, I actively participated in this project and contributed to its development. Over two years of fruitful collaboration, we forged a strong partnership. In 2019, during one of our visits, I proposed the idea for my PhD project, which focused on assessing temperature impacts in the city of São Paulo, a subject that had been raised in several occasions as a growing concern for the city. The idea was well-received, marking the beginning of a new chapter in our partnership.

Throughout this journey, my primary liaison at USP has been Prof. Dr. Adelaide Nardocci, an Associate Professor of Environmental Risk Assessment and Risk Management at the Department of Environmental Health within the School of Public Health at USP. Prof. Nardocci's expertise in quantitative risk assessment of environmental contaminants, vulnerability assessment, and spatial analysis has played a pivotal role in ensuring its relevance and applicability to the context of São Paulo. Initially, I had planned to make periodic visits to São Paulo and possibly stay for a longer duration to work at USP. The purpose was to gain firsthand familiarity with the city's context and specific aspects. However, due to the impact of COVID-19, these plans became unfeasible. Consequently, the input and support from Prof. Nardocci became even more crucial. She played a significant role in facilitating negotiations with data providers and guiding me through the bureaucratic processes of obtaining ethical approvals and data requests.

More information about USP School of Public Health can be found in their website.

6.7.3 Research visit at the European Space Agency ϕ-lab

I am a firm believer in the power of collaborations and multidisciplinary efforts in research. As a result, I made the decision to apply for a placement at the prestigious ESA ϕ-lab. The mission of ESA's ϕ-lab is incredibly inspiring: to accelerate the future of Earth Observation through transformational innovations. These innovations have the potential to completely revolutionize industries or create entirely new ones by harnessing cutting-edge technologies. The ultimate goal is to enhance the competitiveness of the European Earth Observation industrial and research sectors on a global scale. At ϕ-lab, they manage the entire innovation pipeline. This involves generating or discovering

groundbreaking ideas, carefully selecting and nurturing the most promising ones, and then developing the corresponding solutions or products through rigorous testing and refinement. The dedication to bringing these innovations to market adoption is evident in their commitment to investing in the best ideas through the InCubed program.

Following discussions with Dr. Rochelle Schneider, the health link at the ϕ-lab, I successfully secured a 3.5-month research placement at the headquarters of the ESA's ϕ-lab in Frascati, Italy. Commencing in August, this opportunity allowed me to engage with the team and dedicate the subsequent three months to deepening and expanding my knowledge of machine learning and . I had the opportunity to delve deeply into various models and methodologies. Among the models I extensively learned and explored were RF, XGboost, and Long short-term memory (LSTM) networks. Additionally, I gained valuable insights into several other methodologies, including cross-validation approaches, feature selection methods, and hyperparameter tuning approaches. Moreover, I familiarized myself with AOA under the guidance of Hanna Meyer. In terms of technical skills, I honed my proficiency in Python coding language and utilized the Google Collab platform for my work. Furthermore, I had exposure to ESA's satellites fleet and gained hands-on experience working with hyper-spectral and multi-spectral satellite products. As part of my learning journey, I attended the ESA-ECMWF workshop, which provided me with valuable networking opportunities and further enriched my understanding of the field. This culminated with the development of a machine learning pipeline to predicting temperature with high spatiotemporal resolution in São Paulo, which was presented in Chapter 4 of this thesis.

Finally, my stay at ESA's ϕ-lab has led to exciting opportunities for potential collaborations. One significant achievement was successfully securing, together with two colleagues, a Symposium on remote sensing for health applications at the 35^{th} Conference in Taiwan, which is scheduled for September 2023. In this Symposium, where I will serve as the co-chair and moderator, researchers from around the globe will share their insights on the opportunities and challenges of Earth Observation in epidemiological research. Moreover, my time at ESA's ϕ-lab provided a platform for engaging in discussions on various future research. One of these involved further developments on the AOA and CV approach, in collaboration with Hanna Meyer. Additionally, I explored the potential for collaboration with Prof Mattia Crespi from Sapienza University in Rome on the use of LST. Beyond academic collaborations, my interactions during the stay extended to stakeholders and policymakers, including a visit from the Irish minister who showed particular interest in my project. This engagement with policymakers highlights the real-world impact and relevance of my work. Furthermore, I had the opportunity to connect with industry-related companies like Ororatech, a thermal intelligence start-up, who was interested in my views on how epidemiologist could use their data and invited me to present my work at their company.

Additional information about ESA's -lab can be found here.

6.8 Tables

Table 6.1: **Individual determinants re-coded.** Table with the original codes for each category of individual determinant, their description (in Brazilian Portuguese) and their re-codification.

Group	Original code	Description	Re-coded to...
Occupation[a]	0101-0312	Membros das Forcas Armadas, Polic e Bombeiros	Comercial
	1111-1499	Membros Superiores do Poder Publico	Office
	2011-2650	Profissionais das ciencias e artes	Office
	3001-3960	Tecnicos de nivel medio	-
	4101-4250	Trabalhadores servicos administrativos	Office
	5100-5250	Trab serv, Vendedores comercio e lojas	Comercial
	6100-6450	Trab agropec, florestais e pescaa	$\text{Ind}_a gr$
	7100-8659	Trab da producao de bens e serv industriais	$\text{Ind}_a gr$
	9100-9922	Trab em servicos de reparacao e manutencao	-
	0989 / 9999 / NA	missing	NA
Years of education	1	None	¡3 years
	2	Between 1 and 3 years	¡3 years
	3	Between 4 and 7 years	Between 4 and 11 years
	4	Between 8 and 11 years	Between 4 and 11 years
	5	More than 12 years	¿12 years
	0 / 9 / NA	missing	NA
Ethnic group	1	Branca (whites)	White
	2	Preta (Blacks)	Mixed
	3	Amarela (Asians)	Asians
	4	Parda (mixed race)	Mixed
	5	Indigena (Indigena)	-
	0 / 9 / NA	missing	NA
Marital status	1	Solteiro	Single
	2	Casado	Married
	3	Viúvo	Widow
	4	Separado judicialmente/divorciado	Single
	5	União estável	Married
	0 / 9 / NA	Unknown / incorrect / missing	NA
Sex	1	Male	Male
	2	Female	Female
	3	Unknown	NA
	0 / 9 / NA	Unknown / incorrect / missing	NA
Age	Years as digits (x)	xx	"0-4", "5-9",. . . , "80-84", "+85"

Table 6.2: **Temperature statistics by contextual determinant.** Summary statistics of temperature (5^{th}, 50^{th} and 95^{th} percentiles) for each contextual determinant category.

Determinant	Class	5^{th}	50^{th}	95^{th}	Determinant	Class	5^{th}	50^{th}	95^{th}
idhm	1	14.11	20.06	24.95	legalemplQ	1	14.56	20.58	25.56
	2	14.36	20.33	25.21		2	14.6	20.6	25.54
	3	14.51	20.5	25.42		3	14.57	20.57	25.53
	4	14.62	20.63	25.58		4	14.56	20.56	25.47
	5	14.59	20.6	25.58		5	14.3	20.25	25.15
idhm_edu	1	14.16	20.15	25.1	minwage1Q	1	14.59	20.61	25.59
	2	14.34	20.29	25.17		2	14.62	20.63	25.59
	3	14.52	20.52	25.43		3	14.53	20.52	25.46
	4	14.61	20.62	25.56		4	14.5	20.51	25.42
	5	14.58	20.59	25.57		5	14.34	20.3	25.19
idhm_inc	1	14.18	20.11	25.02	minwage2Q	1	14.57	20.59	25.58
	2	14.4	20.37	25.21		2	14.61	20.62	25.58
	3	14.48	20.48	25.41		3	14.64	20.63	25.54
	4	14.64	20.64	25.59		4	14.47	20.46	25.41
	5	14.59	20.6	25.58		5	14.31	20.26	25.15
idhm_lon	1	14.18	20.11	25.02	unemplQ	1	14.6	20.61	25.61
	2	14.4	20.37	25.2		2	14.63	20.64	25.57
	3	14.47	20.47	25.4		3	14.51	20.51	25.47
	4	14.64	20.64	25.59		4	14.44	20.42	25.34
	5	14.59	20.6	25.58		5	14.4	20.36	25.26
IPVS_V10	0	14.29	20.19	25.04	empl_inc_ineqQ	1	14.28	20.24	25.14
	1	14.55	20.57	25.56		2	14.51	20.49	25.42
	2	14.59	20.6	25.55		3	14.64	20.64	25.54
	3	14.44	20.43	25.34		4	14.63	20.64	25.6
	4	14.43	20.39	25.33		5	14.52	20.55	25.55
	5	14.29	20.26	25.1	giniQ	1	14.42	20.33	25.24
	6	14.34	20.3	25.21		2	14.43	20.44	25.32
black_mixedpop	<10%	14.58	20.6	25.6		3	14.57	20.57	25.57
	inbetween	14.52	20.52	25.44		4	14.56	20.65	25.56
	>60%	14.27	20.23	25.15		5	14.59	20.57	25.57
whitepop	<20%	14.27	20.22	25.15	vuln_houseQ	1	14.6	20.6	25.57
	inbetween	14.52	20.53	25.46		2	14.45	20.48	25.42
	>90%	14.6	20.6	25.59		3	14.47	20.45	25.37
otherpop	0%	14.49	20.47	25.31		4	14.47	20.44	25.39
	inbetween	14.51	20.51	25.46		5	14.58	20.58	25.49
	>30%	14.61	20.63	25.67	house_povQ	1	14.57	20.6	25.59
elderly	<5%	14.26	20.21	25.07		2	14.63	20.63	25.59
	>15%	14.6	20.61	25.62		3	14.58	20.58	25.54
	inbetween	14.57	20.58	25.53		4	14.46	20.46	25.37
build_matQ	1	14.7	20.58	25.39		5	14.34	20.29	25.16
	2	14.43	20.54	25.59	indoor_manulaboQ	1	14.45	20.49	25.47
	3	14.57	20.58	25.56		2	14.33	20.33	25.33
	4	14.52	20.52	25.43		3	14.54	20.54	25.42
	5	14.37	20.35	25.28		4	14.62	20.61	25.48
crowdQ	1	14.53	20.56	25.55		5	14.64	20.6	25.53
	2	14.6	20.61	25.58	indoorQ	1	14.53	20.49	25.46
	3	14.6	20.6	25.54		2	14.57	20.56	25.4
	4	14.47	20.46	25.36		3	14.41	20.42	25.34
	5	14.38	20.34	25.23		4	14.49	20.5	25.48
waterQ	1	14.4	20.36	25.29		5	14.57	20.58	25.56
	2	14.51	20.52	25.46	outdoor_manulaboQ	1	14.67	20.66	25.61
	3	14.52	20.53	25.49		2	14.64	20.64	25.61
	4	14.57	20.56	25.47		3	14.55	20.57	25.51
	5	14.56	20.58	25.53		4	14.48	20.49	25.4
educQ	1	14.28	20.25	25.18		5	14.26	20.21	25.12
	2	14.53	20.52	25.4	rururb	RURAL	13.78	19.53	24.38
	3	14.58	20.59	25.51		URBANO	14.51	20.52	25.45
	4	14.61	20.61	25.57	zonas	MCQUA	14.18	20.12	24.99
	5	14.57	20.6	25.59		MCUUS	13.3	18.68	23.72
illiterQ	1	14.58	20.6	25.57		MEM	14.67	20.66	25.6
	2	14.63	20.62	25.62		MPEN	13.94	19.96	24.58
	3	14.6	20.6	25.52		MQU	14.64	20.64	25.59
	4	14.46	20.46	25.37		MRVU	14.56	20.56	25.53
	5	14.32	20.28	25.18		MRVURA	14.12	20.08	24.9
						MUC	14.58	20.6	25.6
					inf_settl	formal	14.55	20.56	25.51
						informal	14.34	20.31	25.18

Table 6.3: **Temperature statistics by individual determinant.** Summary statistics of temperature (5^{th}, 50^{th} and 95^{th} percentiles) for each individual determinant category.

Individual determinant	**Categories**	$\mathbf{5}^{th}$	$\mathbf{50}^{th}$	$\mathbf{95}^{th}$
Age group	50-54	14.60	20.63	25.42
	55-59	14.54	20.55	25.43
	60-64	14.53	20.54	25.43
	65-69	14.50	20.52	25.42
	70-74	14.50	20.47	25.42
	75-79	14.52	20.51	25.45
	80+	14.46	20.46	25.49
Years of education	<3 years	14.43	20.44	25.37
	4-11 years	14.49	20.48	25.43
	>11 years	14.55	20.57	25.51
Marital status	Single	14.51	20.52	25.42
	Married	14.50	20.50	25.44
	Widow	14.50	20.49	25.48
Ethnic group	White	14.53	20.52	25.47
	Mixed	14.45	20.46	25.36
	Asian	14.56	20.54	25.57
Occupation	commercial	14.49	20.49	25.42
	ind_agr	14.45	20.45	25.40
	office	14.58	20.61	25.55
Sex	Females	14.53	20.53	25.47
	Males	14.49	20.50	25.43

6.9 Figures

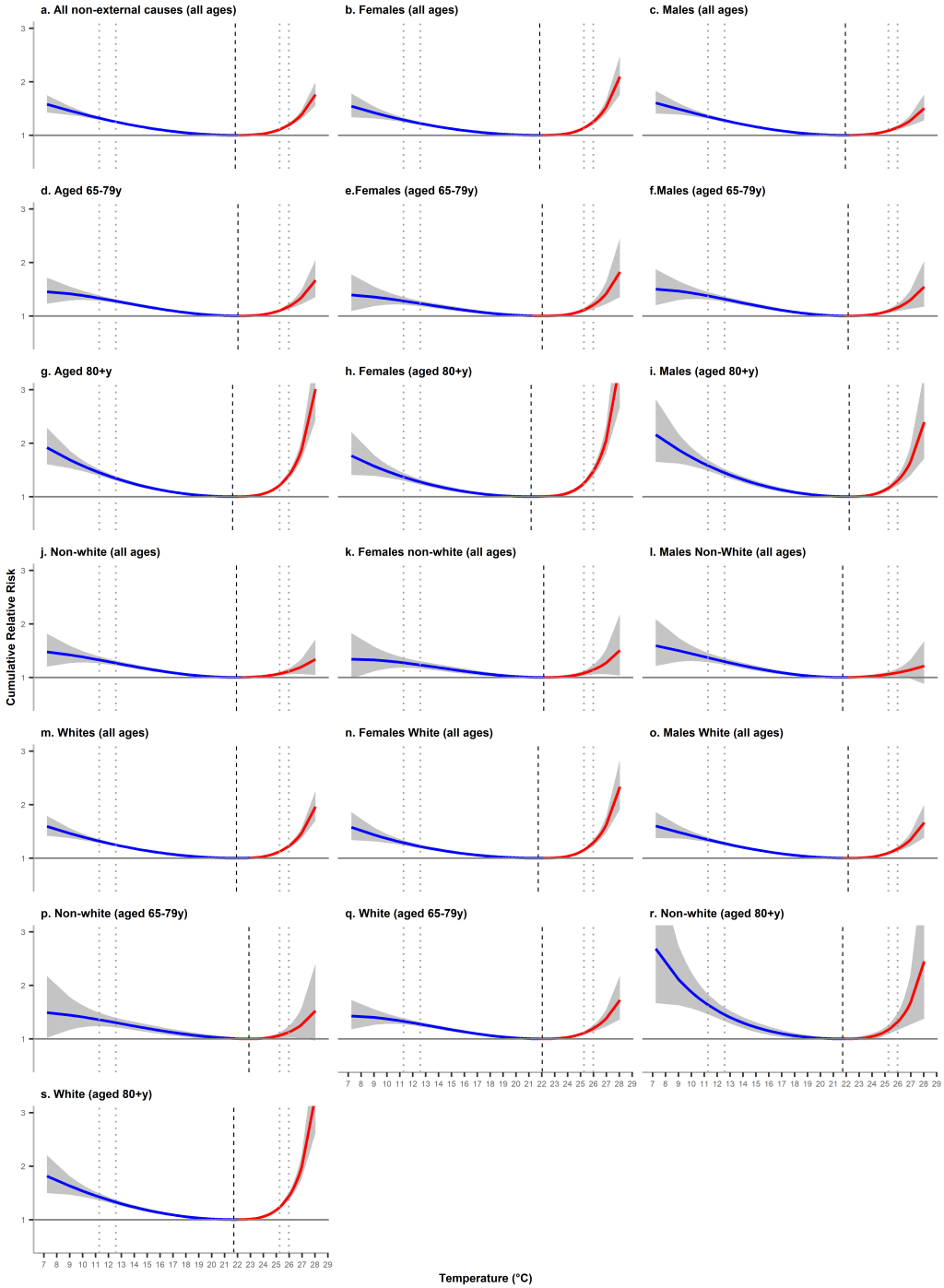


Figure 6.2: **Cumulative temperature-mortality association for 2000-18 by population group.** Overall cumulative relative risk (cRR, solid line) and 95% CI by age, gender and ethnic group. In blue, the risk to cold temperature and in red to heat. The MMT (dashed), and the 1^{st},$10^{t}h$, $90^{t}h$, and $99^{t}h$ percentiles of the temperature distribution (dotted) are shown as vertical lines. The cRR curves were re-scaled to the group-specific overall MMT.

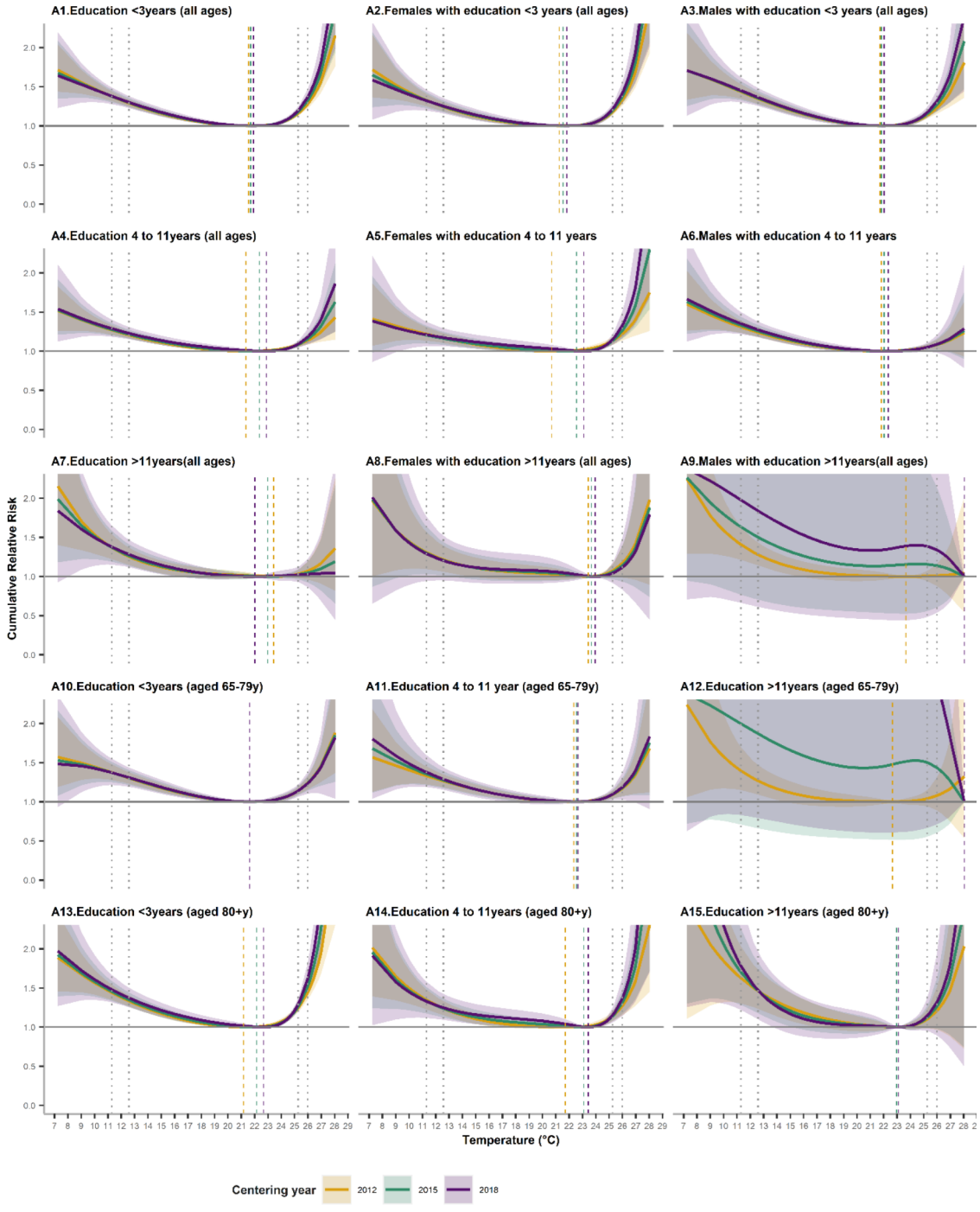


Figure 6.3: **Temperature-mortality association by education group,** for 2011, 2015 and 2018 with 95%CI. Cumulative relative risk (solid line) and 95%CI (shaded) for the temperature-mortality association for 2011, 2015, and 2018, for each education level category. The MMT (dashed), the $1^{st}/10^{th}$, and $90^{th}/99^{th}$(dotted) percentiles of the temperature distribution are shown as vertical lines

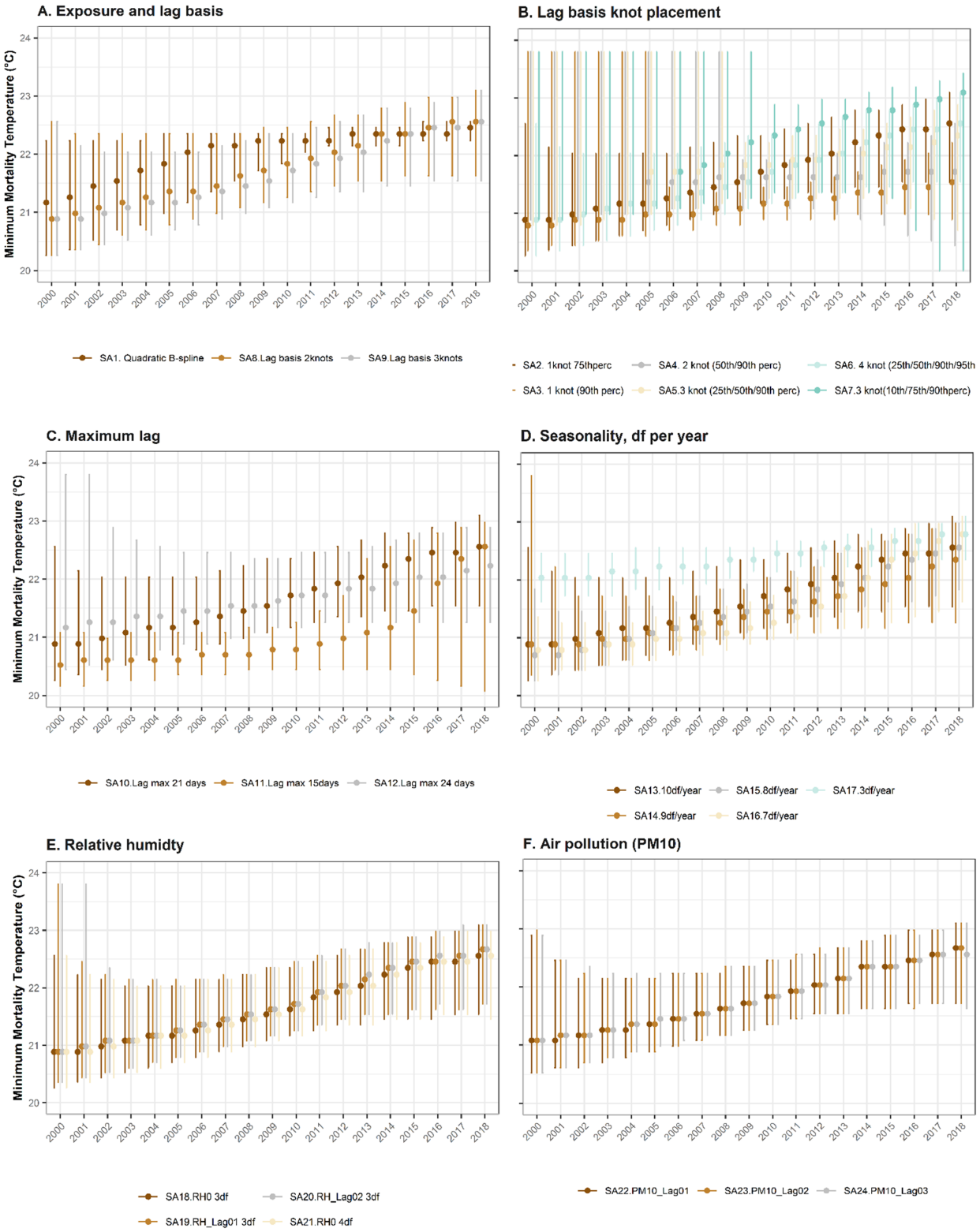


Figure 6.4: **Trends of MMT by sensitivity analyses.** Annual MMT estimates and 95%CI by varying modelling (A to D) choices and controlling for relative humidity, RH (E) and PM_{10} (F). Computed on all-non external causes only.

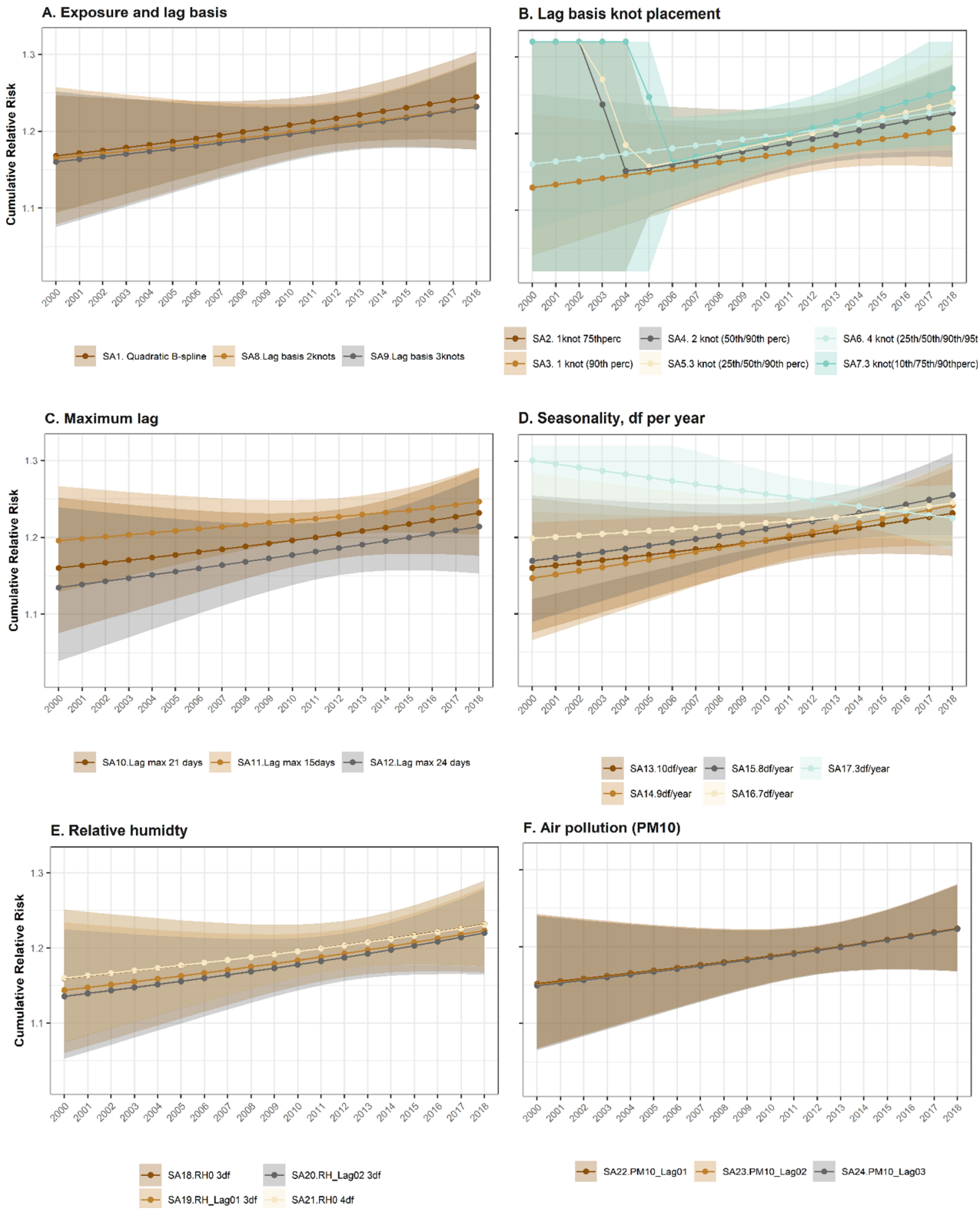


Figure 6.5: **Trends of the cRR for extreme heat[a] by sensitivity analysis.** The cRR and 95%CI are provided by varying modelling (A to D) choices and controlling for relative humidity (E) and PM10 (F). Computed on all-non external causes only. [a]Extreme heat: cRR at the 99^{th} percentile vs. MMT

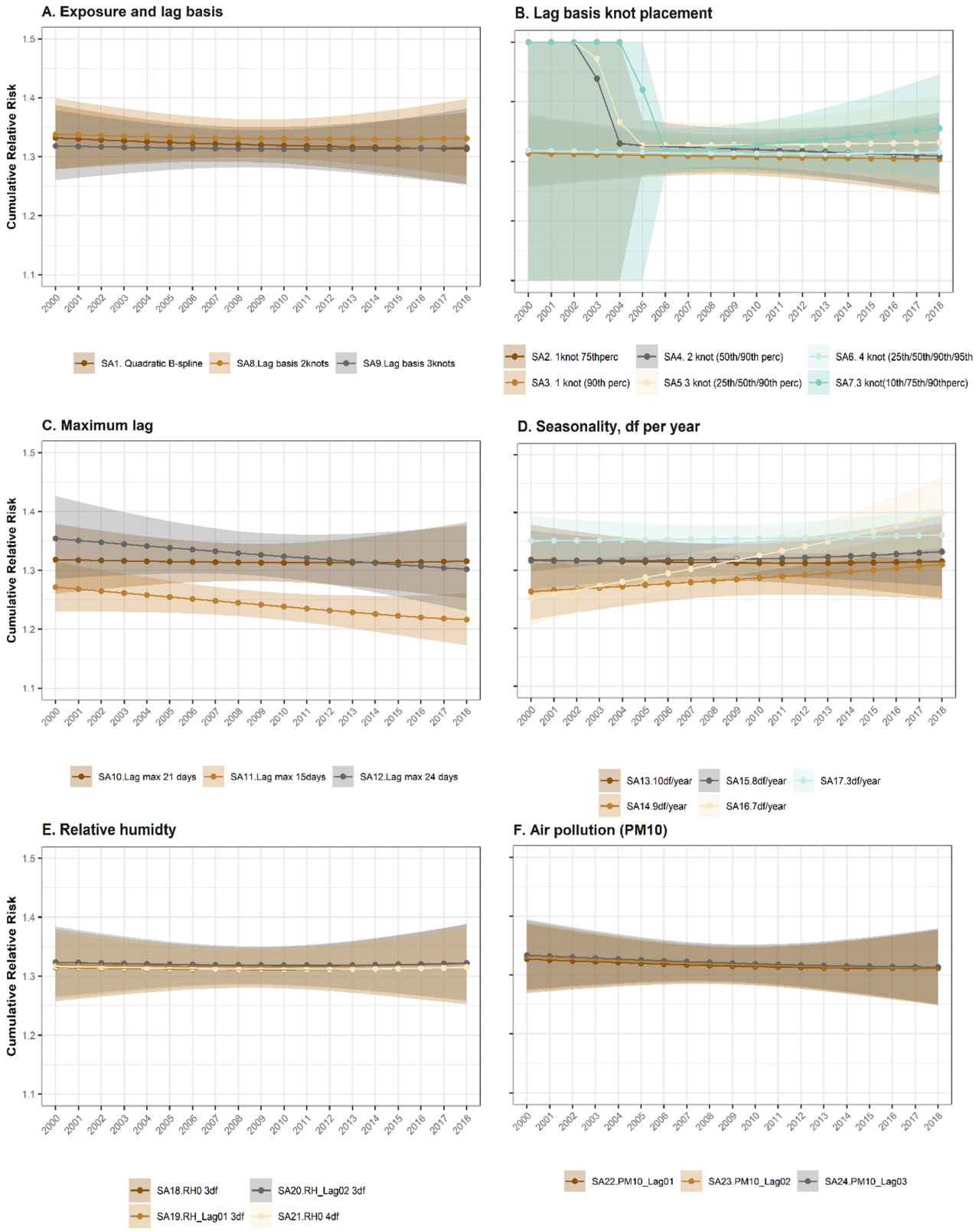


Figure 6.6: **Trends of the cRR for extreme cold[a] by sensitivity analysis.** The cRR and 95%CI are provided by varying modelling (A to D) choices and controlling for relative humidity (E) and PM10 (F). Computed on all-non external causes only. [a]Extreme cold: cRR at the 1^{st} percentile vs. MMT

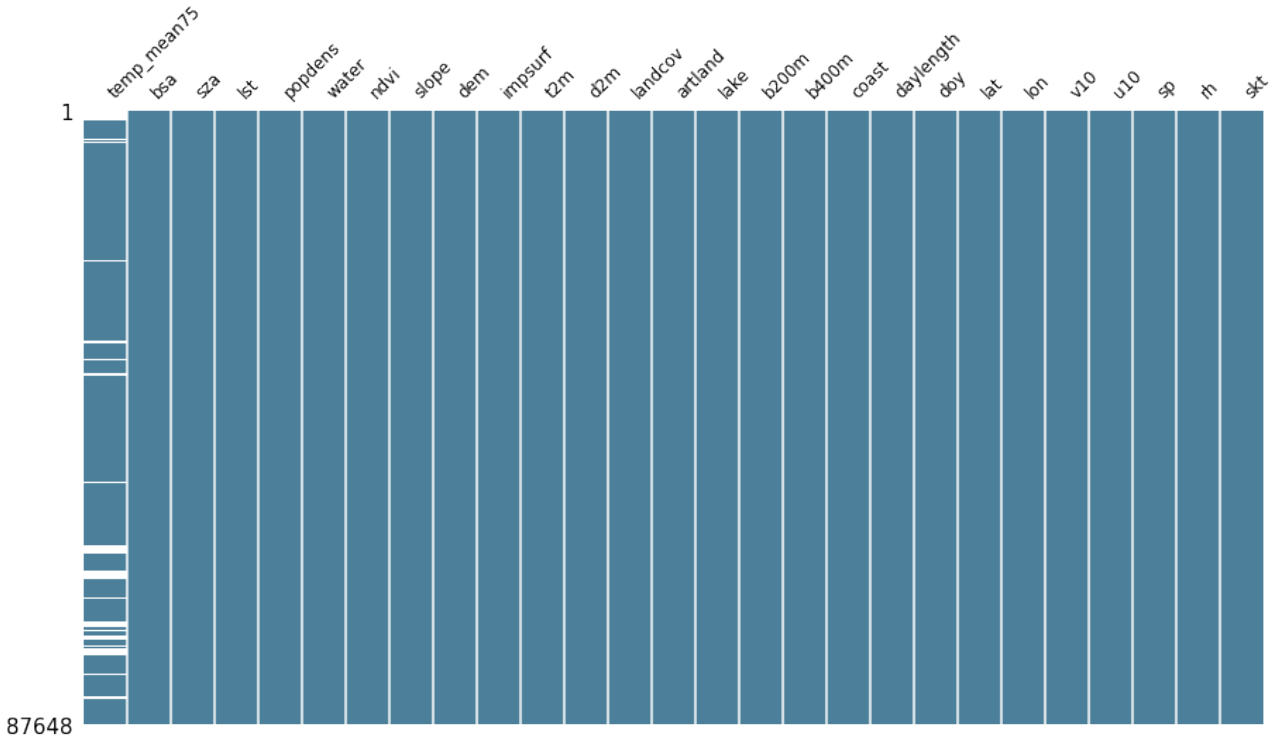


Figure 6.7: **Missing values.** Matrix plot showing patters of missing values (white) as opposed to complete records (steel blue) for the target variable and each input feature.

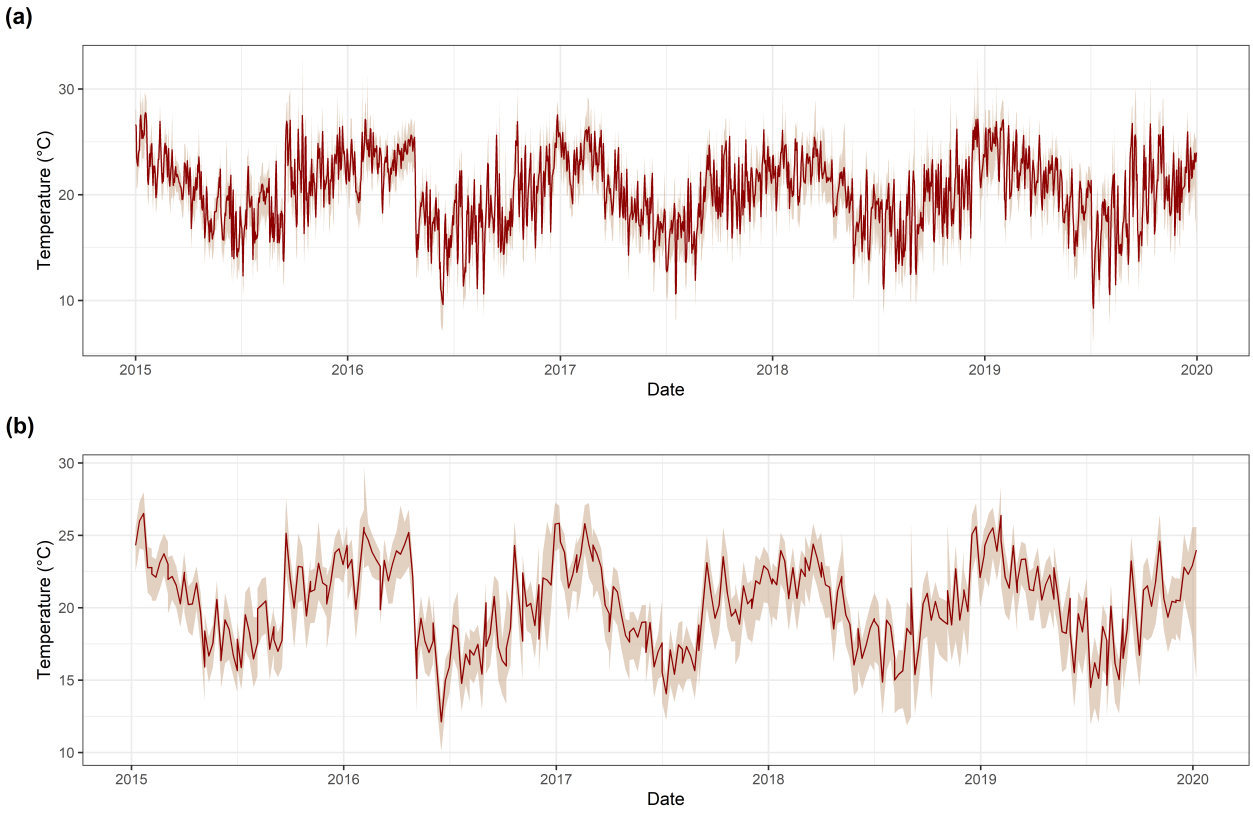


Figure 6.8: **Daily and weekly and temperature from included meteorological stations.** Average (solid line) and minimum/maximum (shaded) mean temperature by (a) day and (b) week between 2015 and 2020.

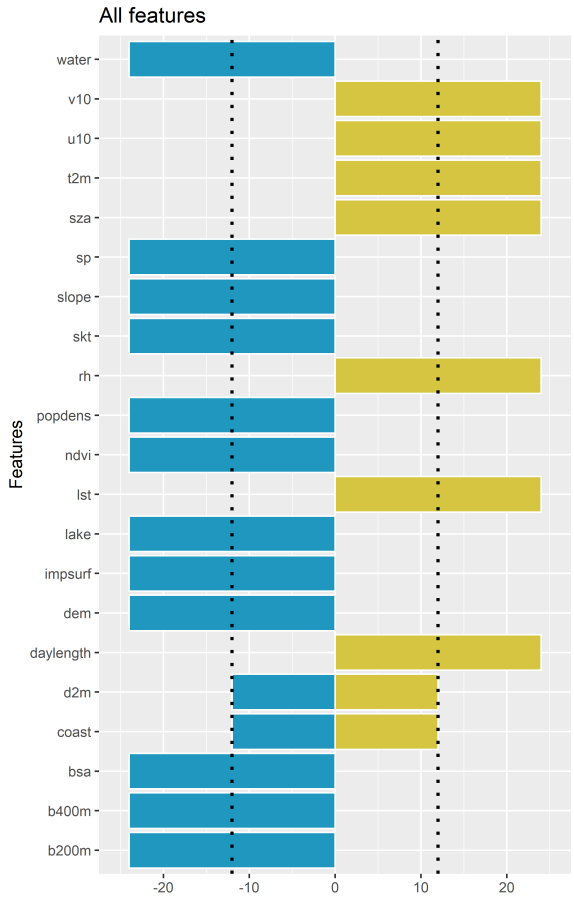


Figure 6.9: **Feature selection agreement by hyper-parameter tuning run.** Score in favour of including (ochre) or excluding (blue) a feature according to the forward feature selection algorithm. The score is calculated as the number of agreeing models out of the total hyperparameter combinations tested (n=24). The dashed vertical line indicates the median score (-12 and 12). Performance assessed using spatial cross-validation.

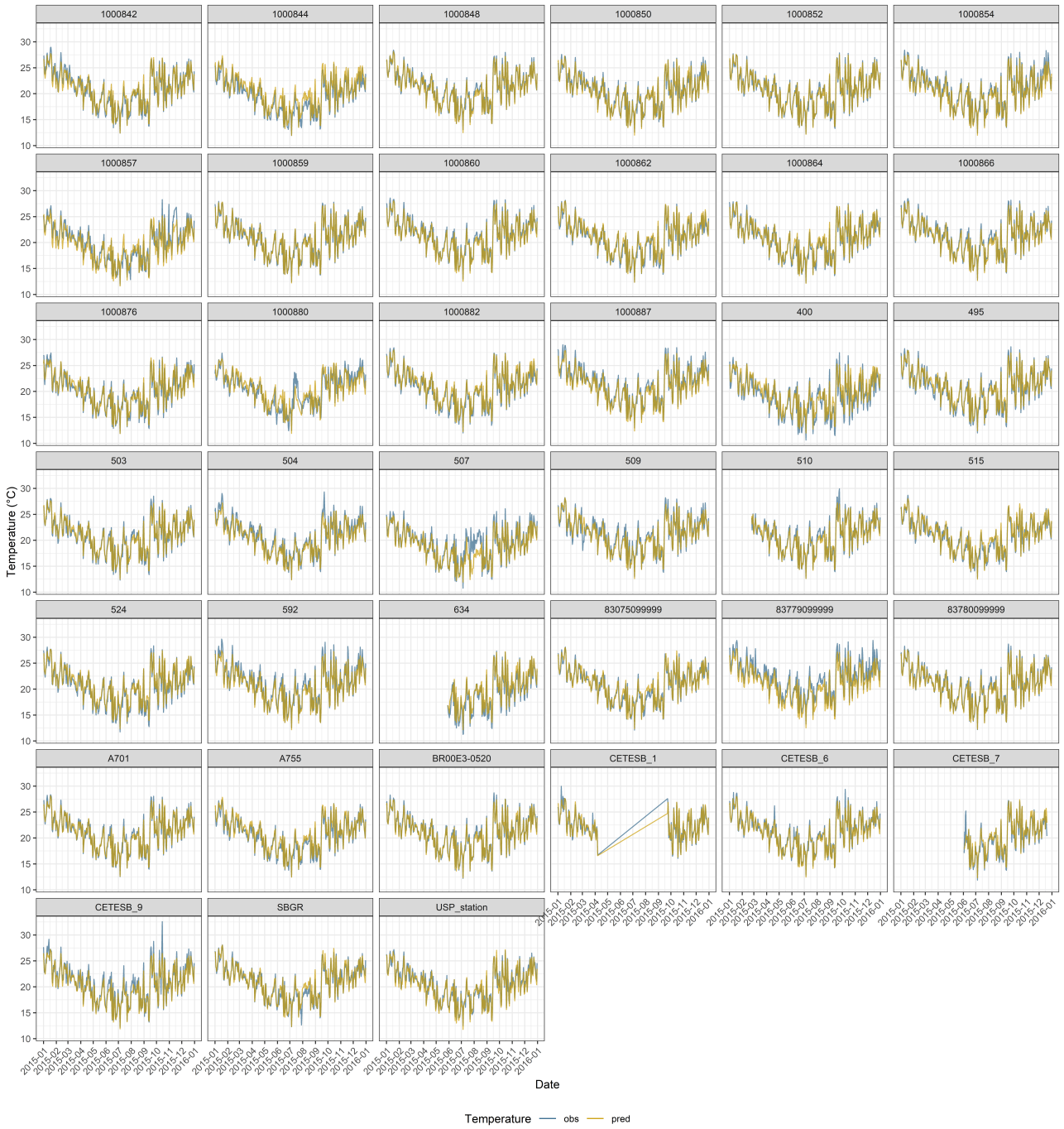


Figure 6.10: **Observed and predicted temperature at each station in 2015.** Predictions from the RF algorithm calculated through the spatial-CV (ochre) and observations (blue) for the each station for 2015 (randomly selected year). Note: CETESB_1 had a gap with no data; ignore the connecting lines.

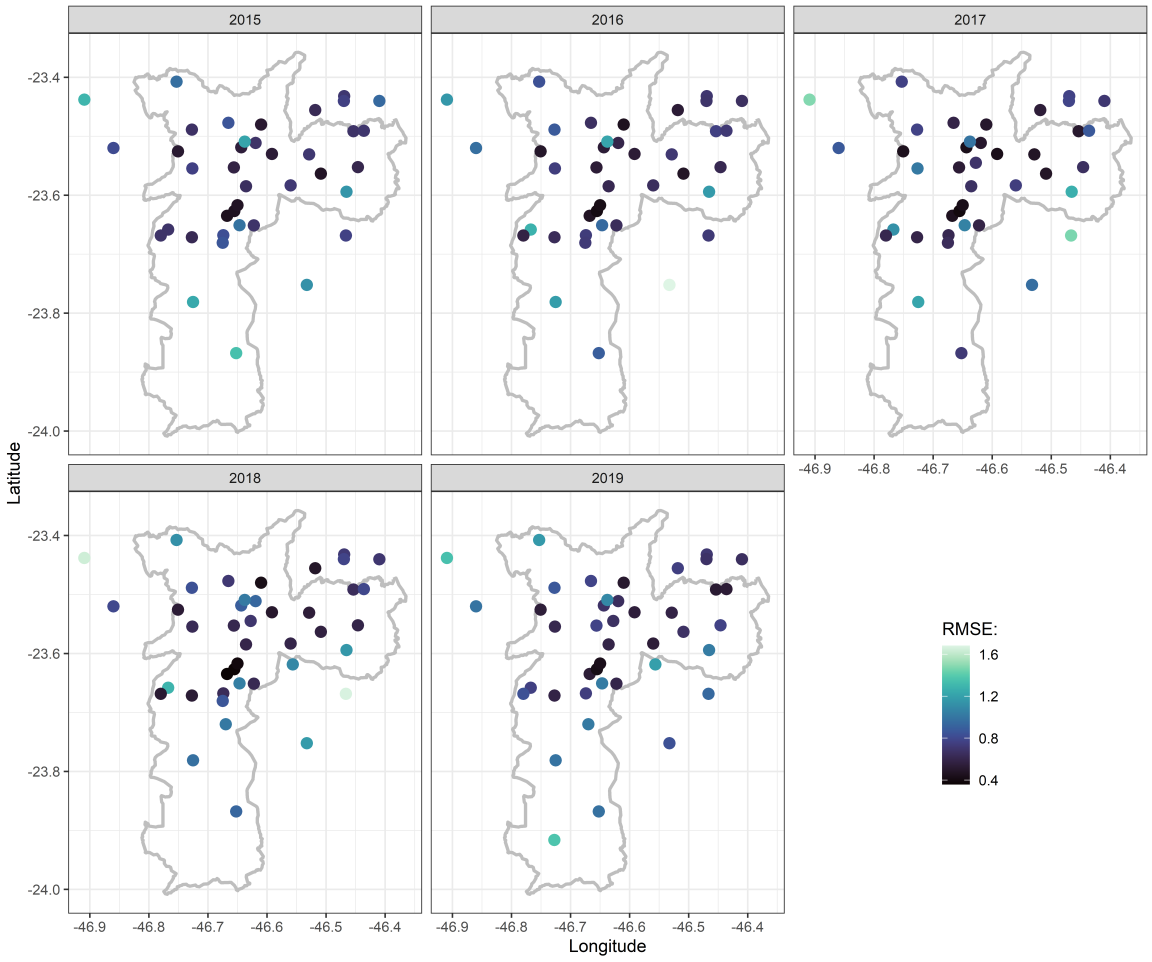


Figure 6.11: **RMSE per station for each year.** Average RMSE (°C) per station by year, assessed through spatial-CV.

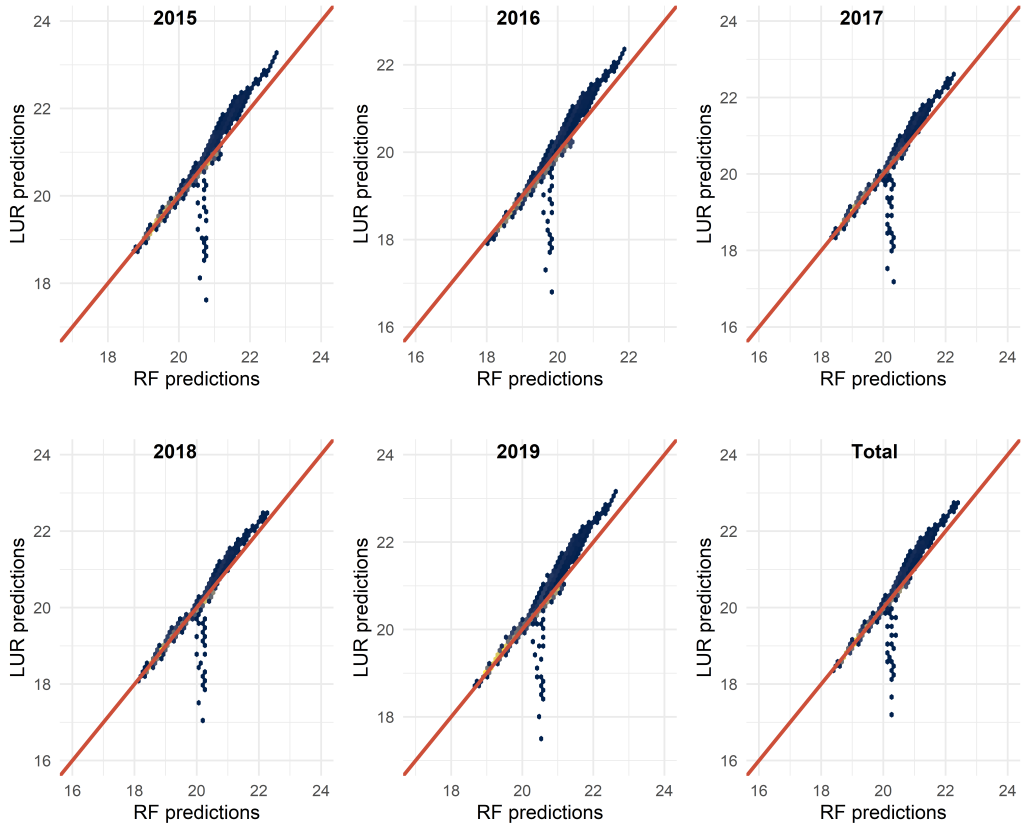
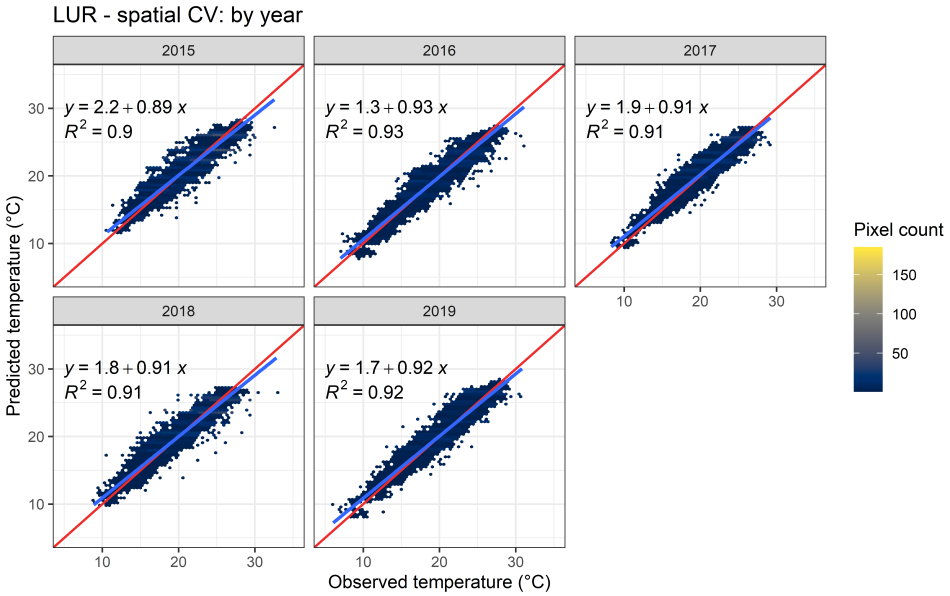
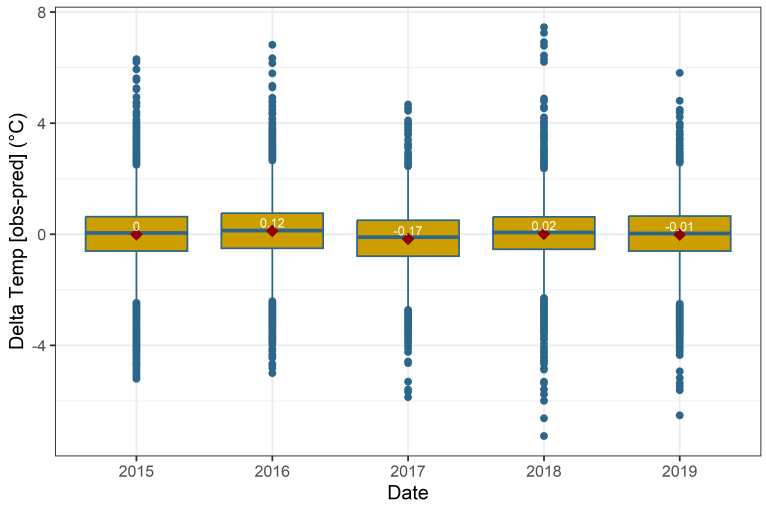


Figure 6.12: **Density scatter plot of RF and LUR predicted temperatures overall and by year.** Density scatter plots of the annual average daily mean temperature predicted by the RF (reference) and LUR models. The red and blue lines represent the 1:1 line and the linear regression, respectively. The R^2 and regression equation are shown for each plot.

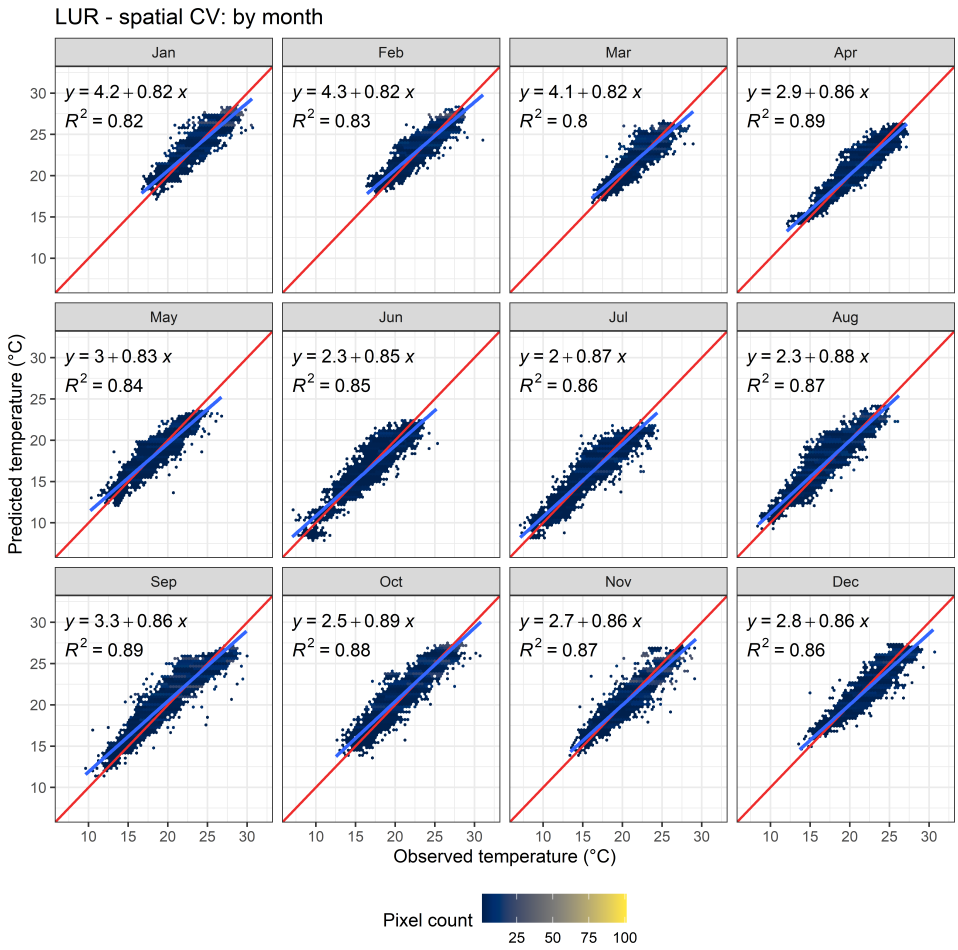


(a) Scatter plots for 10-fold spatial CV

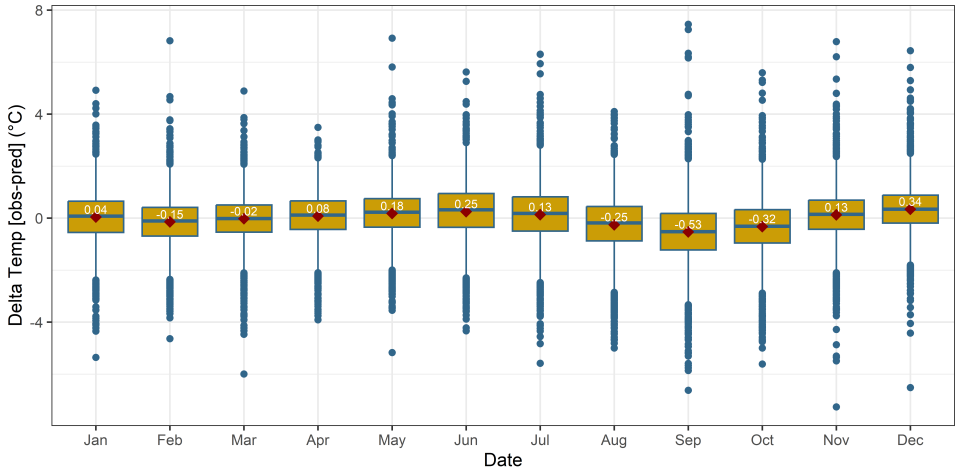


(b) $\Delta Temp$ box plot

Figure 6.13: **Annual differences in model accuracy and $\Delta Temp$ for LUR model.**(a)M Annual density scatter plots of LUR predicted daily mean temperature versus daily mean temperature assessed through spatial-CV. The red and blue lines represent the 1:1 line and the linear regression, respectively. The R^2 and regression equation are shown for each plot. (b) Boxplot showing the difference between observed and predicted daily temperatures across all stations for each month in the study period (2015-2019). The red diamond indicates the mean; value label above.



(a) Scatter plots for 10-fold spatial CV



(b) $\Delta Temp$ box plot

Figure 6.14: **Monthly differences in model accuracy and $\Delta Temp$ for LUR model.**(a)Monthly density scatter plots of LUR predicted daily mean temperature versus daily mean temperature assessed through spatial CV. The red and blue lines represent the 1:1 line and the linear regression, respectively. The R^2 and regression equation are shown for each plot.(b) Boxplot showing the difference between observed and predicted daily temperatures across all stations for each year in the study period (2015-2019). The red diamond indicates the mean; value label above.

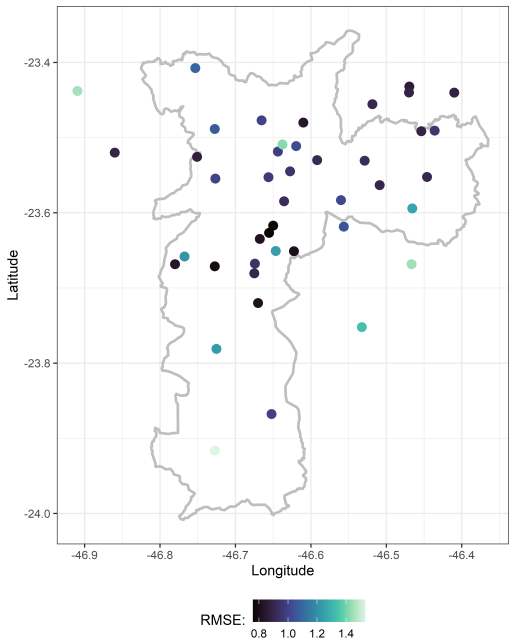
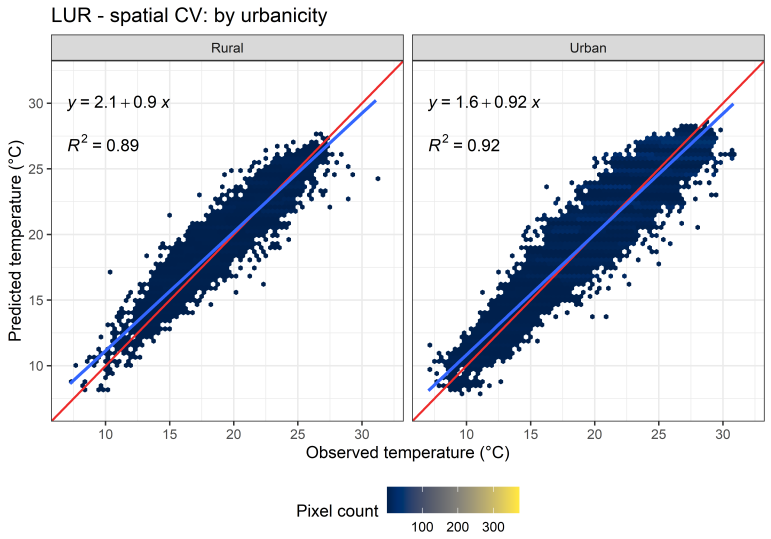
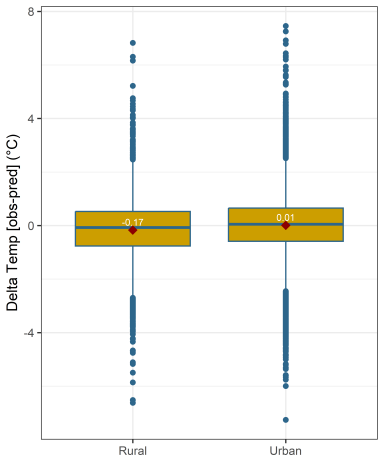


Figure 6.15: **RMSE per station for the period 2015-2019.** Average RMSE (°C) per station the period 2015-2019, assessed through spatial-CV.



(a) Scatter plots for 10-fold spatial CV



(b) $\Delta Temp$ box plot

Figure 6.16: **Rural/urban differences in model accuracy and $\Delta Temp$ for LUR model.**(a) Density scatter plots of LUR predicted daily mean temperature versus daily mean temperature for stations classified as urban Vs. rural, assessed through spatial-CV. The red and blue lines represent the 1:1 line and the linear regression, respectively. The R^2 and regression equation are shown for each plot.(b) Boxplot showing the difference between observed and predicted daily temperatures ($\Delta Temp$) across all stations by urban/rural classification across all study period (2015-2019). The red diamond indicates the mean; value label above.

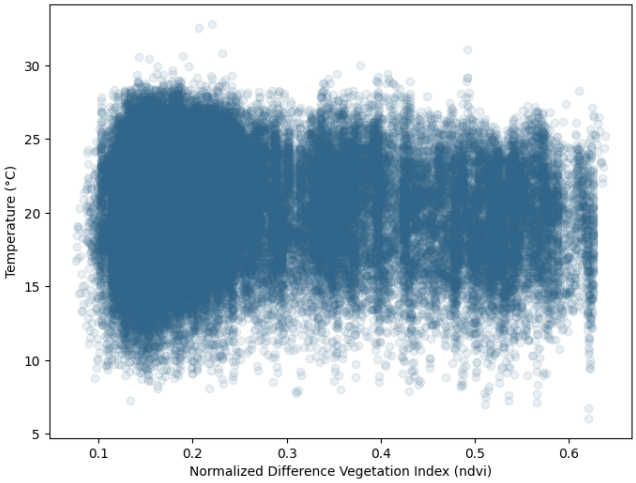


Figure 6.17: **Correlation plot between temperature and NDVI scores.** Recorded temperature at each meteorological station (y axis) and NDVI score (x axis)

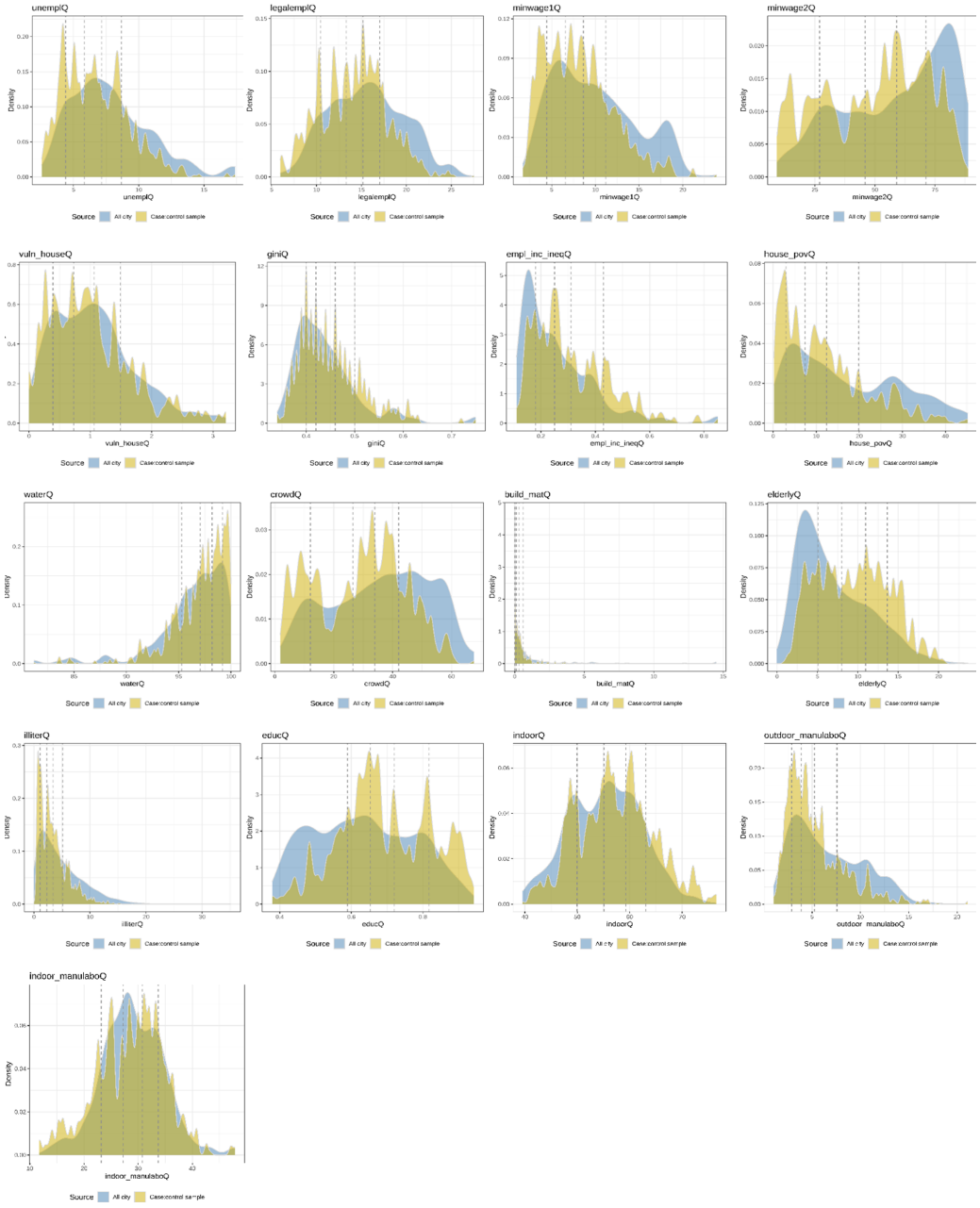


Figure 6.18: **Histogram of population and sample contextual determinants (continuous).** Distribution of each variable across all HDUs in the study area (blue) and in my case/control sample (ochre), and the corresponding quintiles (dashed vertical lines). Only continuous variables displayed.

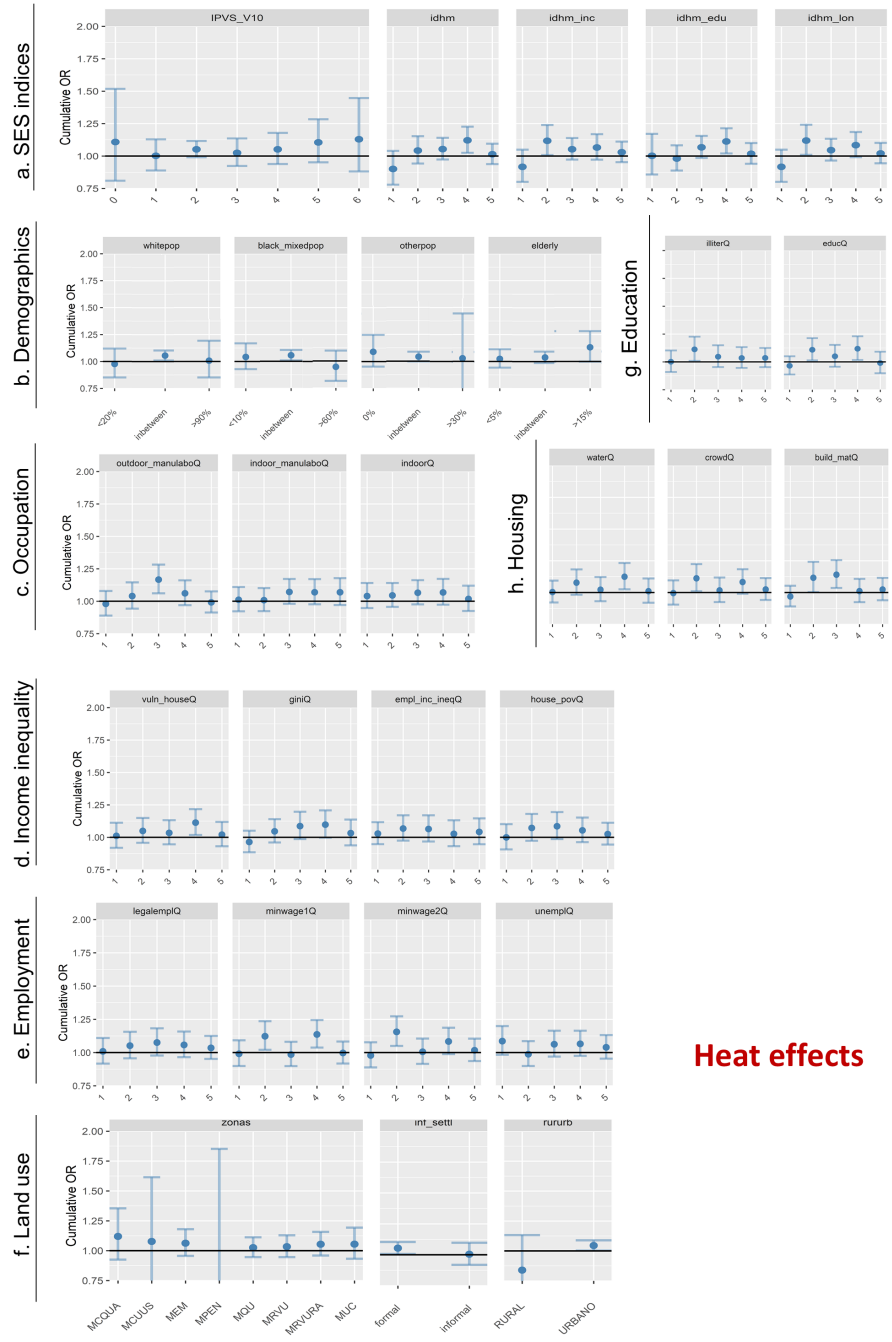


Figure 6.19: **cOR associated to heat by contextual determinant quintiles.** Cumulative Odds Ratio (cOR) (points) and 95% confidence intervals(bars) for non-external cause mortality associated to heat in relation to the the overall MMT (23.29°C), stratified over contextual determinants. Estimates accumulated over 0-24 lag days.

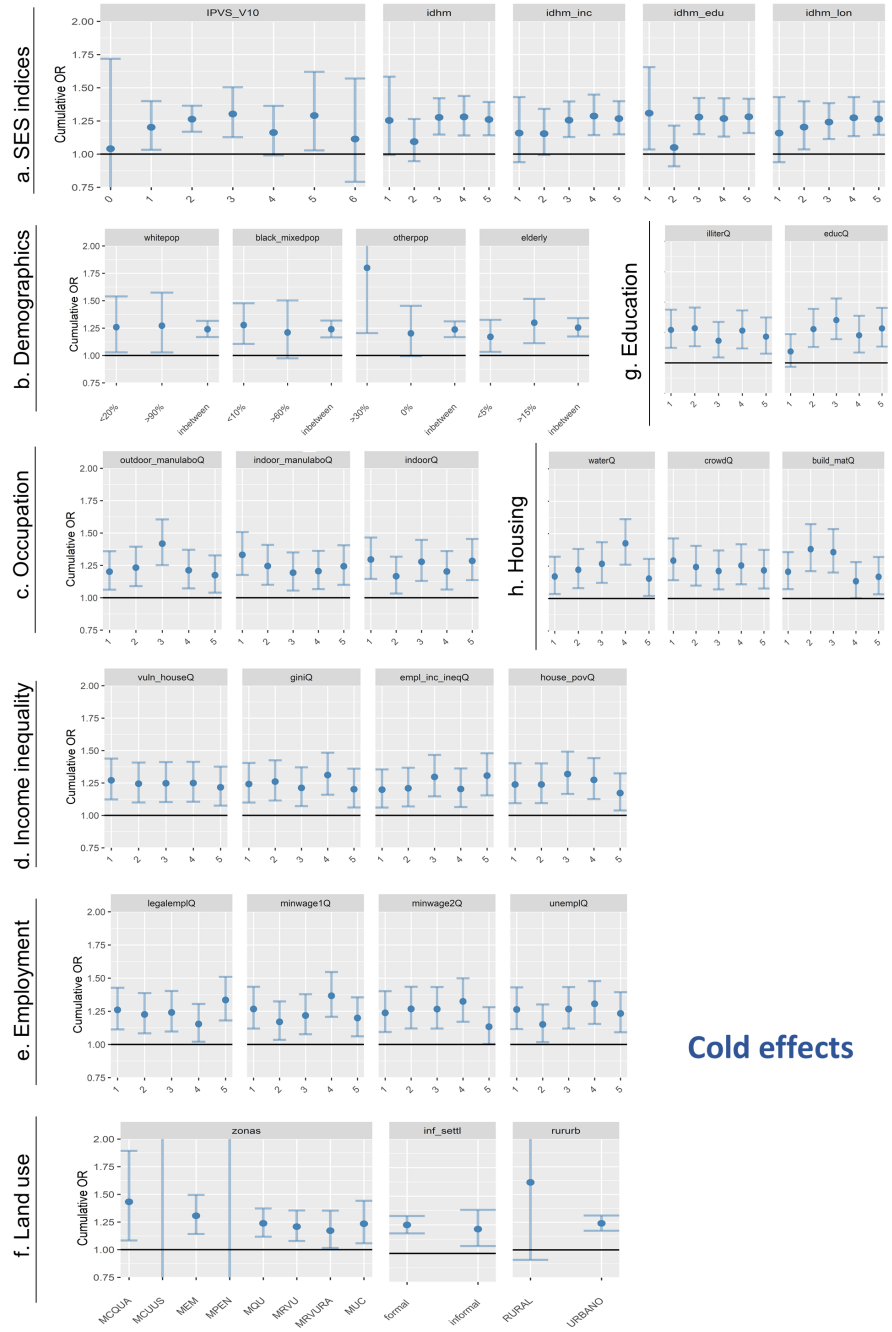


Figure 6.20: **cOR associated to cold by contextual determinant quintiles.** Cumulative Odds Ratio (cOR) (points) and 95% confidence intervals(bars) for non-external cause mortality associated to cold in relation to the the overall MMT (23.29°C), stratified over contextual determinants. Estimates accumulated over 0-24 lag days.