

IMPERIAL

IMPERIAL COLLEGE LONDON

DEPARTMENT OF COMPUTING

Graph-Based Spatiotemporal Machine Learning: From Methods to Cardiac Fibrillation

Alexander Jenkins

March 18, 2026

Submitted in part fulfilment of the requirements for the degree of Doctor of Philosophy in
Computing of Imperial College London and the Diploma of Imperial College London.

*To my parents,
my sister Embeth,
and my girlfriend Alice*

my structure in the chaos

Abstract

Signals evolving jointly across space and time arise in systems as diverse as climate dynamics, traffic networks, and cardiac electrophysiology. When the spatial domain is irregular, graphs provide a natural abstraction, with nodes representing locations and edges encoding their relationships. Graph-based spatiotemporal machine learning combines graph signal processing and graph neural networks with temporal sequence models to learn representations of signals evolving over time and topology. Yet fundamental challenges persist: observations are often sparse and incomplete, safety-critical predictions demand reliable uncertainty, non-stationarity shifts the governing distributions, and the graph topology itself is frequently unknown or time-varying and must be inferred.

These challenges are acute in atrial fibrillation, the most common sustained cardiac arrhythmia, where chaotic activation propagates across the irregular geometry of the atrium. Clinical mapping records signals at sparse catheter locations, yielding incomplete observations on a graph-structured domain; beat-to-beat variability prevents these non-contemporaneous recordings from being assembled into a coherent global picture. Reconstructing global dynamics, quantifying their reliability, and identifying the conduction pathways sustaining the arrhythmia are each prerequisites for moving beyond empirical ablation toward personalised, mechanism-directed therapy.

This thesis develops new methodologies for graph-based spatiotemporal machine learning, forming a progressive *reconstruct–analyse–characterise* pipeline. FIBMAP formulates cardiac mapping as spatiotemporal graph signal reconstruction, using a graph recurrent neural network to recover global dynamics from sparse measurements, validated on optical mapping and large-scale clinical recordings. COREL introduces a distribution-free conformal prediction framework that exploits relational structure among correlated time series to provide calibrated, adaptive uncertainty quantification without modifying the base predictor. ADACGP develops a sparsity-aware algorithm for learning time-varying graph topologies from streaming data at constant per-iteration cost. While FIBMAP is task-specific, COREL and ADACGP extend to any domain requiring analysis of time-varying signals on irregular network structures.

Contents

Abstract	i
List of Figures	vii
List of Tables	viii
List of Acronyms	x
Declaration of Originality	xi
Copyright	xii
Acknowledgements	xiii
1 Introduction	1
1.1 Motivation and Relevance	1
1.2 Research Questions	6
1.3 Main Contributions	8
1.4 Thesis Organisation	11
1.5 List of Publications	12
I Background	16
2 Theoretical Background	17
2.1 Signals and Systems on Regular Domains: Time Series Models	17
2.1.1 Linear Time-Invariant Systems and the Shift Operator	18
2.1.2 Random Signals and Stationarity	21
2.1.3 Autoregressive Models	21
2.1.4 Adaptive Filtering	23
2.1.5 Neural Sequence Models	25
2.2 Signals and Systems on Irregular Domains: Graph Signal Processing	27
2.2.1 Graphs and Signals	28
2.2.2 The Graph Shift Operator	29

2.2.3	Graph Filters	31
2.2.4	Spectral Graph Theory	32
2.2.5	Graph Machine Learning	35
2.3	Time-Varying Signals and Systems on Irregular Domains: Graph-Based Spatiotemporal Models	39
2.3.1	Classical Spatiotemporal Models on Graphs	40
2.3.2	Architectures for Spatiotemporal Graph Neural Networks	42
2.3.3	Beyond Global Models, Fixed Topologies and Point Predictions	46
3	Cardiac Electrophysiology and Fibrillation	49
3.1	Electrical Activity of the Heart	49
3.2	Electrical Propagation Dynamics	52
3.3	Mechanisms of Atrial Fibrillation	54
3.4	Treatment of Atrial Fibrillation	57
3.5	Electrophysiological Mapping	58
3.5.1	Clinical Invasive Mapping	59
3.5.2	Optical Mapping	62
3.6	Summary and Research Questions	63
II	Imputation and Reconstruction	65
4	Improving Diffusion Models for ECG Imputation	66
4.1	Introduction	67
4.2	Preliminaries	70
4.2.1	ECG Imputation	70
4.2.2	Diffusion Models for Time Series Imputation	72
4.3	Methodology	73
4.3.1	Problem Formulation	73
4.3.2	PULSEDIFF: Template-Guided Diffusion for ECG Imputation	75
4.4	Experiments	80
4.4.1	Experimental Setup	80
4.4.2	Results	82
4.4.3	Ablation Study	84
4.5	Discussion	87
4.6	Summary and Relation to Thesis	89

5	Learning to Reconstruct Global Atrial Fibrillation Dynamics From Sparse Measurements	92
5.1	Introduction	93
5.2	Preliminaries	95
5.3	Methodology	95
5.3.1	Problem Formulation	96
5.3.2	FibMap: Architecture	96
5.3.3	FibMap: Optimisation	101
5.4	Experiments	102
5.4.1	Datasets	102
5.4.2	Experimental Setup	106
5.5	Results	113
5.5.1	Rat VF: Optical Mapping Validation	113
5.5.2	Human AF: Non-Contact Mapping Validation	116
5.5.3	Translational Showcase: Contact Mapping with HD Grid	121
5.6	Discussion	124
5.7	Summary and Relation to Thesis	128
III	Analysis and Characterisation of Spatiotemporal Systems	131
6	Relational Conformal Prediction for Calibrated Uncertainty Quantification	132
6.1	Introduction	133
6.2	Preliminaries	135
6.2.1	Problem Formulation	135
6.2.2	Conformal Prediction for Time Series	136
6.3	Methodology	138
6.3.1	Relational Quantile Predictor	139
6.3.2	Learning the Relational Structure	142
6.3.3	Theoretical Analysis	143
6.3.4	Adaptation	144
6.4	Experiments	145
6.4.1	Time Series Forecasting Benchmarks	146
6.4.2	Controlled Synthetic Environment	149
6.4.3	Adaptation Experiments	151
6.4.4	Prediction Interval Width Optimisation	152

6.5	Discussion	153
6.6	Summary and Relation to Thesis	154
7	Online Graph Topology Learning via Time-Vertex Adaptive Filters	156
7.1	Introduction	157
7.2	Methodology	158
7.2.1	Problem Formulation	158
7.2.2	Objectives for Online Optimisation	159
7.2.3	Online Estimation Algorithm	161
7.3	Experiments	168
7.3.1	Controlled Synthetic Environment	169
7.3.2	Assessing the Structure and Stability of Cardiac Fibrillation	180
7.4	Discussion	184
7.5	Summary and Relation to Thesis	186
8	Conclusion and Future Directions	188
8.1	Conclusion	188
8.2	Future Directions	190
8.2.1	Methodological Extensions	190
8.2.2	Clinical Translation	193
8.3	Closing Remarks	195
	Bibliography	196

List of Figures

1.1	Euclidean versus graph spatial domains.	2
1.2	Spatiotemporal signals on irregular domains.	3
1.3	Atrial fibrillation mapping formulated as a graph-based spatiotemporal learning problem.	6
1.4	Thesis structure and conceptual overview.	15
2.1	Illustration of discrete-time signals.	18
2.2	Illustration of the time-shift operator.	20
2.3	Scalar and vector autoregressive models.	23
2.4	Signal domains as graphs of increasing structural irregularity.	28
2.5	Graph shift operator on regular and irregular topologies.	30
2.6	Polynomial graph filter of order $K=2$ applied to an impulse, $\mathbf{x} = \boldsymbol{\delta}_0$, on a 5-node graph topology.	32
2.7	Single-layer update at node i under convolutional and message-passing aggregation from neighbours $(j_1, \dots, j_4 \in \mathcal{N}_i)$	37
2.8	The ‘light cone’ of the causal graph process model.	41
3.1	Cardiac electrical activity across spatial scales.	51
3.2	Comparison of normal sinus rhythm and atrial fibrillation.	55
3.3	The electrophenotype spectrum of cardiac fibrillation.	56
3.4	Overview of catheter ablation for atrial fibrillation.	57
3.5	The two paradigms of clinical invasive mapping.	60
3.6	Optical mapping of cardiac fibrillation.	63
4.1	The proposed PULSEDIFF methodology.	69
4.2	Comparison of imputation results on ECG waveforms from the PTB-XL dataset.	71
4.3	Overview of the PULSEDIFF conditioning mechanism.	79
4.4	Reconstruction accuracy of probabilistic diffusion models.	85
4.5	Performance of probabilistic diffusion models on classification and beat detection downstream tasks.	86
5.1	FIBMAP inputs and architecture.	97

5.2	Training and fine-tuning workflow for FIBMAP.	109
5.3	Fine-tuning convergence on validation hearts/patients.	110
5.4	Qualitative reconstruction of rat ventricular fibrillation from simulated sparse sequential contact mapping.	115
5.5	Qualitative reconstruction of human AF from simulated sparse sequential contact mapping.	117
5.6	Sensitivity analysis of FIBMAP reconstruction performance.	120
5.7	FIBMAP’s learned state and embedding spaces.	122
5.8	Proof-of-concept clinical translation of FIBMAP to real-world HD Grid contact mapping.	123
5.9	Representative FIBMAP reconstructions from HD Grid contact mapping.	125
6.1	Overview of the COREL framework.	141
6.2	Comparison between (a) the true graph and (b) the graph structure learned by COREL in GPVAR.	151
7.1	Convergence of ADACGP variants on the K-Regular topology.	171
7.2	Estimated graph shift operator matrices across four graph topologies.	175
7.3	Performance of ADACGP variants for different graph topologies.	177
7.4	Computational complexity analysis of graph topology learning algorithms.	178
7.5	Local stability analysis of fibrillation dynamics.	182
7.6	Global analysis of fibrillation dynamics across varying rotigaptide concentrations.	183

List of Tables

3.1	Comparison of electrophysiological mapping modalities used in this thesis.	59
4.1	Imputation performance on the PTB-XL dataset with 30% transient missing values.	83
4.2	Imputation performance on the PTB-XL dataset with 30% extended missing values.	84
4.3	Ablation study for PULSEDIFF components.	87
5.1	Characteristics of patients in the AcQMap dataset.	105
5.2	Hyperparameter values tested during FIBMAP training.	108
5.3	Inter-observer agreement for phase singularity detection.	111
5.4	Quantitative reconstruction performance on rat VF optical mapping test set.	114
5.5	Phase singularity detection performance on rat VF optical mapping.	116
5.6	Quantitative reconstruction performance on human AF AcQMap test set.	118
5.7	Phase singularity detection performance on human AF AcQMap test set across precision thresholds.	119
6.1	Uncertainty quantification performance for $\alpha = 0.1$ across three datasets and three base predictors.	148
6.2	Uncertainty quantification performance on the synthetic GPVAR dataset.	149
6.3	Effect of adaptive embedding updates on COREL performance for the CER-E dataset ($\alpha = 0.1$).	151
6.4	Effect of prediction interval width optimisation.	152
7.1	Comparison of graph topology estimation algorithms for $N = 50$	173
7.2	Scaling of per iteration time and memory complexity.	179

List of Acronyms

AF atrial fibrillation. 4–7, 9, 10, 54–58, 61, 62, 64, 93–97, 103–107, 111–113, 116, 118–121, 123–128, 157, 187, 189, 190, 193–195

AR autoregressive. 21, 22, 25

CGP causal graph process. 41, 46, 48, 157–159, 166–169, 186

DDPM denoising diffusion probabilistic model. 27, 68, 72, 73, 87, 89, 129

DSP digital signal processing. 18, 19, 21, 29

ECG electrocardiogram. 9, 11, 51, 52, 55, 66–68, 70, 72, 73, 75, 76, 80, 81, 87–90, 93, 188

ERP effective refractory period. 53

FIR finite impulse response. 20, 22, 31, 32

GCN graph convolution network. 36, 37

GFT graph Fourier transform. 21, 33, 34, 36

GNN graph neural network. 3, 35, 36, 38, 39, 130

GRNN graph recurrent neural network. 44, 94–96, 98, 99, 101, 124, 126, 128, 129

GRU gated recurrent unit. 26, 44, 98, 139–141, 146

GSP graph signal processing. 2, 158, 186

IGFT inverse graph Fourier transform. 34

IIR infinite impulse response. 20, 22, 26

LMS least mean square. 24, 25, 144, 158, 159, 161, 168, 172

LSTM long short-term memory. 26, 44, 45

LTI linear time-invariant. 19–21

MA moving average. 22, 42

MAE mean absolute error. 109, 110, 114, 118, 119, 127, 146

MF matrix factorisation. 95, 109, 117, 126

mHealth mobile health. 67, 70, 73, 80, 89

MLP multilayer perceptron. 38, 39, 43, 97, 139, 140

MPNN message-passing neural network. 38, 98, 99, 129

MRE mean relative error. 110, 114, 118

MSE mean squared error. 23, 80, 82, 85, 88, 110, 114, 118, 161, 166

PS phase singularity. 110, 111, 114, 118

RLS recursive least squares. 24, 25, 158, 159, 168, 172

RNN recurrent neural network. 2, 26, 46, 70, 95, 110, 146, 149, 150

SCP split conformal prediction. 136, 137, 139, 147–149

STGNN spatiotemporal graph neural network. 3, 40, 42, 43, 46, 47, 95, 96, 128, 141, 146, 147, 149–151, 155, 157

STMP spatiotemporal message-passing. 43–45

STT space-then-time. 46

T&S time-and-space. 44–46, 95, 101, 126, 128

TTS time-then-space. 45, 46, 110, 126, 139, 140, 144–146

VAR vector autoregressive. 22, 39–42, 46, 48, 150, 158, 172, 181, 184

VF ventricular fibrillation. 56, 95, 103, 104, 107, 108, 113, 114, 124, 158, 180–182, 184, 185, 194

WSS wide-sense stationary. 21, 40

Declaration of Originality

I, Alexander Jenkins, hereby declare that this thesis, and the research to which it refers, are the product of my own work. Any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline. A full list of references is given in the bibliography. All figures are my own unless otherwise stated and referenced. I confirm that this work has not been submitted, in whole or in part, for any other degree or professional qualification at this or any other university. Regarding co-authored work, my specific contributions to any collaborative research included in this thesis are explicitly defined in the introduction.

Copyright

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International Licence (CC BY NC-SA).

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, do not use it for commercial purposes and share any derivative works under the same licence.

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 those 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.

© 2026 Alexander Jenkins

Acknowledgements

First I would like to thank my supervisors, Professor Fu Siong Ng and Professor Danilo Mandic. It has been the greatest pleasure to spend these years working with you. I have deeply valued your mentorship, support, patience, and all that you have taught me. The PhD has been a great experience intellectually, and I sincerely thank you for the opportunity to pursue this research training under your guidance. I would also like to thank my close collaborator Professor Cesare Alippi for warmly welcoming me to his team at the Università della Svizzera italiana as a Visiting Researcher in 2024 and 2025, and for his continued inspiration and encouragement.

Thank you to all my colleagues and friends. You have made my research journey and PhD journey truly special. I'd like to first thank my close friends who I have been so fortunate to know for many years. Thank you for always keeping me smiling and laughing: James Mills, Tom Mullen, Alex Dunne, Davie Morgan, Dafydd Matheson, Joseph Gilson, Ieuan Mercer, Jordan Bowen, Jon Bishop, Alex Thomas, Mike Gray, Sam Hudson, Lewis Prole, and Mitch Ayres.

To my friends and colleagues at Imperial, to whom there are so many, you have shaped every aspect of this journey with many memorable moments, discussions, and the occasional 5-pint pitchers. To name just a few. From the Bio & Algo groups: Matteo Bermond, Edoardo Occhipinti, Michael Thornton, Yao Xu, Kriton Konstantinidis, Zehua Chen, Pauline Bourigault, Marek Zylinski, Annie Chiu, Yutong Zheng, Harry Davies, and Amir Nassibi. From the CardiacEP.AI team at the National Heart and Lung Institute: Joseph Barker, Ahmed El-Medany, Vineesh Kappadan, Nikesh Bajaj, Huiyi Wu, Tony Hua, Aidan Birdi, Konstantinos Patlatzoglou, and Gulrukh Khattak. And of course, to my great swimming and sauna friend, Simon Hanassab. It has been a pleasure.

Thank you to my colleagues and great friends at the Graph Machine Learning Group (GMLG), Lugano, Switzerland. First among them, I'd like to thank my friend Andrea Cini. It has been a pleasure working so closely with you during my PhD, and I am truly grateful for both your collaboration and your friendship. From the Bernina Rossa to St Patrick's Day in Liverpool, thank you to the GMLG team for all the brilliant memories: Andrea Cini, Alessandro Manenti, Tommaso Marzi, Giovanni De Felice, Valentina Moretti, Ivan Marisca, Luca Butera, Daniele Zambon, Miguel Nogales, Giovanni Donghi, Federico Bombardieri, and of course, the Chief Assistant Officer, Steve Imoscopi. Thank you also to other great people and friends I met and made during this time, particularly, Peter Kaeser and Laurence Kaeser, for such a warm welcome to Collina d'Oro.

I would also like to thank Britta Ross, Ahmed Fetit, Safa Abu-Sway, Aldo Faisal, and the UKRI CDT in AI for Healthcare at Imperial for all your support. The CDT has been a fantastic initiative, without whose support this work would not be possible (Grant No. P/S023283/1). It has been an honour to be part of it and contribute at the intersection of AI and healthcare.

Finally, I would like to thank my parents, my sister Embeth, and the love of my life, Alice. It is hard for me to put into words just how much your love and encouragement have meant to me. Thank you for always being there, always listening, always offering guidance, and always challenging me to be the best version of myself. I would not have been able to do this without you. This is as much your achievement as it is mine.

Chapter 1

Introduction

1.1 Motivation and Relevance

Signals that evolve in both space and time are among the most complex phenomena encountered in science and engineering. These spatiotemporal signals often exhibit turbulent, chaotic, or non-stationary dynamics, arising in systems as diverse as heat diffusion in materials, electromagnetic wave propagation, and fluid flow. While many such signals evolve over regular Euclidean domains amenable to classical analysis, a substantial class evolves over irregular spatial structures: traffic flows through road networks, energy demand propagates across power grids, weather patterns develop across sensor networks, and electrical waves traverse the curved tissue of the heart. A natural framework for describing such irregular domains is a graph, a mathematical structure consisting of nodes and their pairwise relationships. Figure 1.1 illustrates this distinction for pedestrian flow in an urban environment: while physical space embeds locations in $\mathbb{R}^2$, the relevant structure for modelling movement is the connectivity of walkable paths, where graph distance (640m) diverges substantially from Euclidean distance (470m) due to buildings and other obstacles.

Advances in sensing technology and data infrastructure have dramatically increased the ability to capture spatiotemporal signals at unprecedented resolutions. However, high-dimensional measurements collected continuously across vast sensor networks lead to an explosion in the number of trainable model parameters, rendering statistical learning algorithms prone to overfitting and computational intractability, a phenomenon known as the curse of dimensionality [Bellman, 1961]. These datasets are further characterised by diverse formats, inherent measurement uncertainty, and real-time processing requirements [I. Lee, 2017]. Concurrently, developments in machine and deep learning [LeCun, Bengio, and Hinton, 2015; Bronstein, Bruna, LeCun, et al., 2017], specialised hardware [Reuther et al., 2020], and large-scale optimisation methods [Bottou, Curtis, and Nocedal, 2018] have enabled the training of increasingly powerful models capable of extracting complex patterns from massive datasets. However, the ability to train large models does not obviate the need for careful architectural design. Inductive biases, constraints that prioritise certain solutions over others independent of the observed data, provide a principled ap-

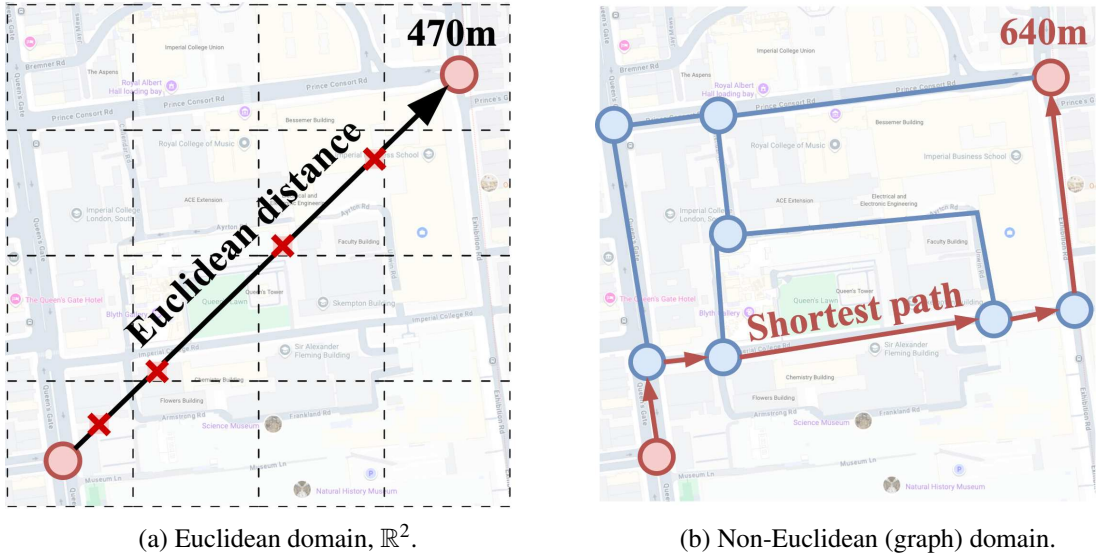
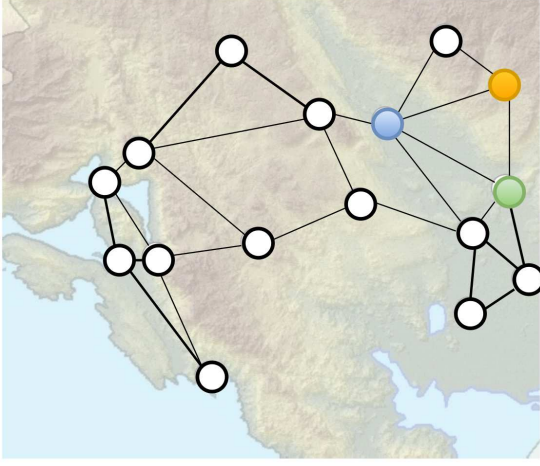


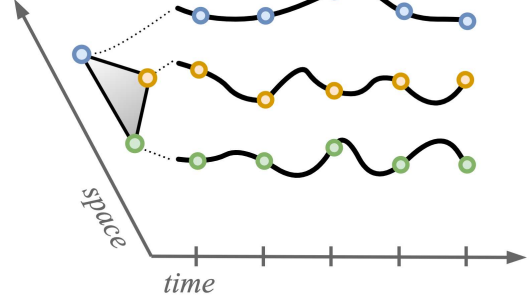
Figure 1.1: Euclidean versus graph spatial domains. (a) Physical space embedded in $\mathbb{R}^2$ with Euclidean distance of 470m between two locations (red crosses indicate inaccessible paths through buildings). (b) The same space represented as a graph where edges denote walkable streets. The shortest path on the graph is 640m, demonstrating that connectivity rather than geometric proximity defines distance for pedestrian flow.

proach to these challenges by trading model flexibility for improved sample efficiency, a trade-off formalised through the bias-variance framework [Geman, Bienenstock, and Doursat, 1992]. Returning to the pedestrian flow example in Figure 1.1, a model that leverages graph connectivity as a *relational inductive bias* [Battaglia et al., 2018] would propagate information along walkable paths rather than through arbitrary spatial relationships, naturally supporting tasks like forecasting crowd movement while dramatically reducing parameter counts.

Modelling spatiotemporal signals requires frameworks that jointly process temporal and spatial domains while exploiting appropriate inductive biases. Classical discrete-time signal processing provides powerful tools for temporal analysis, including linear time-invariant systems, spectral methods, and autoregressive models, treating signals as sequences evolving on regular time grids [Oppenheim, Schafer, and Buck, 1999; Stanković, D. Mandić, et al., 2020b]. The integration of deep learning with signal processing has produced powerful sequence models including recurrent neural networks (RNNs) [Connor, R. D. Martin, and Atlas, 1994; Danilo P Mandić and Chambers, 2001], convolutional neural networks [LeCun, Boser, et al., 1989; Bai, Kolter, and Koltun, 2018], and Transformers [Vaswani et al., 2017; Q. Wen et al., 2023], which excel at processing data on regular lattices such as time series and images. However, many real-world systems exhibit spatial dependencies defined by irregular graph structures rather than regular grids. Graph signal processing (GSP) provides the theoretical foundation for analysing such data by general-



(a) Weather stations form a spatial graph.



(b) Spatiotemporal graph signals.

Figure 1.2: Spatiotemporal signals on irregular domains. (a) Weather stations positioned across terrain form an irregular graph, with edges connecting spatially proximate sensors (adapted from [Stanković, D. Mandić, et al., 2020b]) (b) Temperature measurements at each station (colored nodes from panel (a)) evolve over time, forming spatiotemporal graph signals that require models respecting both spatial connectivity and temporal dynamics.

using notions of frequency, filtering, and convolution to graphs [Shuman et al., 2013; Stanković, D. Mandić, et al., 2020a; Stanković, D. Mandić, et al., 2020b]. Building on this framework, graph neural networks (GNNs) [Scarselli et al., 2008; Bacciu et al., 2020] learn representations that respect graph structure, with spectral approaches operating in the graph frequency domain and spatial approaches propagating information across neighbouring nodes. Spatiotemporal graph neural networks (STGNNs) [B. Yu, H. Yin, and Zhu, 2018; Cini, Marisca, Zambon, et al., 2025] extend these methods by coupling temporal and spatial modules to learn representations of signals evolving over both time and graph topology, enabling tasks such as forecasting, imputation, anomaly detection, and classification [M. Jin et al., 2024]. Figure 1.2 illustrates this framework with weather stations distributed across terrain: the stations form an irregular graph with edges connecting spatially proximate sensors, while temperature measurements at each station evolve over time to form spatiotemporal graph signals [Stanković, D. Mandić, et al., 2020b]. Accounting for relationships between stations, as captured by the graph, is essential for tasks such as forecasting local conditions from network-wide observations.

Alongside the curse of dimensionality in modern big data, spatiotemporal data analysis presents a series of interrelated challenges [Hamdi et al., 2021; Shekhar et al., 2015; Atluri, Karpatne, and V. Kumar, 2018]. Spatiotemporal relationships are complex and implicit, arising from continuous phenomena discretely sampled in space and time [K. V. Rao, Govardhan, and K. C. Rao, 2012], or from data on inherently discrete structures such as road, sensor, and social networks [Shekhar et al., 2015; Atluri, Karpatne, and V. Kumar, 2018]. In both settings, these data violate the inde-

pendence assumptions of classical methods, exhibiting autocorrelation, spatial heterogeneity, and temporal non-stationarity [Shekhar et al., 2015; Miller and J. Han, 2009; Atluri, Karpadne, and V. Kumar, 2018]. Within this landscape, several challenges are particularly pertinent to learning from spatiotemporal graph signals. Real-world data is often incomplete, whether because continuous fields are observed at a finite set of reference points or because network sensors fail and report intermittently [Atluri, Karpadne, and V. Kumar, 2018; Hamdi et al., 2021], necessitating methods for reconstruction that exploit spatial and temporal structure jointly [Shekhar et al., 2015]. Forecasting future system states remains difficult because spatial and temporal dependencies interact in complex, multi-scale ways [Atluri, Karpadne, and V. Kumar, 2018; Sen, H.-F. Yu, and Dhillon, 2019; K. V. Rao, Govardhan, and K. C. Rao, 2012]. Quantifying uncertainty in spatiotemporal predictions is essential for decision-making in safety-critical applications, yet deep spatiotemporal models either provide only point forecasts or rely on probabilistic methods that impose distributional assumptions and significant computational overhead [D. Wu et al., 2021; He and Z. Jiang, 2023], with existing approaches rarely accounting for the dependency structures and non-stationarity inherent in spatiotemporal data [He and Z. Jiang, 2023]. Finally, the graph topology governing spatial relationships is rarely known *a priori* and must be constructed from heuristics or learned from data [M. Jin et al., 2024], a problem compounded by spurious edges from autocorrelation [Atluri, Karpadne, and V. Kumar, 2018] and the recognition that static graphs are inadequate for non-stationary systems [Atluri, Karpadne, and V. Kumar, 2018; Hamdi et al., 2021].

These challenges manifest with particular acuity in cardiac fibrillation, where electrical activation signals evolve over the irregular geometry of cardiac tissue. Atrial fibrillation (AF) is the most common cardiac arrhythmia, with a lifetime risk of one in three and an estimated 37.5 million people affected worldwide [Mou et al., 2018; Lippi, Sanchis-Gomar, and Cervellin, 2021], and is a major cause of stroke, heart failure, and death [Katsanos et al., 2020; Maisel and Stevenson, 2003; P. A. Wolf et al., 1998]. The arrhythmia is characterised by complex spatiotemporal dynamics including rotating spiral waves and transient phase singularities that propagate across the atrial surface [Richard A Gray, Pertsov, and Jalife, 1998; Cheniti et al., 2018]. Catheter ablation, the targeted destruction of tissue that initiates or sustains AF, aims to restore normal rhythm [Kuck et al., 2016], yet success rates remain as low as 50% at five years in persistent AF [Scherr et al., 2015], largely because current strategies fail to tailor interventions to patient-specific mechanisms [F. S. Ng et al., 2020; Deb et al., 2021]. A central barrier is the inability to map AF effectively [Deb et al., 2021; Cheniti et al., 2018; Boyle et al., 2019; Gillis et al., 2013; Roney et al., 2017]. During clinical invasive electrophysiological mapping, a multipolar electrode

catheter is sequentially positioned across the endocardial surface, recording electrical signals as a function of time and position. For organised arrhythmias such as atrial tachycardia, sequential recordings can be computationally stitched into coherent activation maps that precisely localise arrhythmic sources [Özgül et al., 2023]. In AF, however, significant beat-to-beat variation means that sparse, non-contemporaneous measurements cannot be assembled into a reliable picture of the global spatiotemporal dynamics [Roney et al., 2017; Deb et al., 2021; Cheniti et al., 2018].

The data collected during electrophysiological mapping thus provides an incomplete spatiotemporal signal on an irregular spatial domain, precisely the setting for which graph-based methods are designed. First, reconstructing whole-atrial dynamics from limited catheter observations and forecasting the evolution of electrical wavefronts are problems that naturally benefit from spatiotemporal graph learning, where tissue locations form graph nodes and anatomical connectivity defines edges. Second, the high-stakes nature of ablation decisions demands rigorous uncertainty quantification. Finally, the mechanisms sustaining fibrillation can be characterised by learning patient-specific functional connectivity, the preferential pathways of electrical propagation [Handa, X. Li, et al., 2020]. Because AF dynamics exhibit substantial temporal variability in spatiotemporal organisation, with transitions between organised and disorganised states [Everett IV et al., 2001; Ravelli et al., 2007] and non-stationary characteristics [Stridh et al., 2001; M. Sun et al., 2020], models must address non-stationarity through adaptive mechanisms that track evolving connectivity patterns. Figure 1.3 illustrates this formulation, showing how electrophysiological mapping data naturally forms a spatiotemporal graph signal, while outlining the progressive *reconstruct–analyse–characterise* pipeline that structures this thesis.

This thesis seeks to advance both the methodological foundations of graph-based spatiotemporal machine learning and its application to cardiac electrophysiology. Traditional approaches to electrophysiological mapping have predominantly employed conventional signal processing techniques [De Groot et al., 2022] or computational models requiring extensive offline simulation [Boyle et al., 2019]. While graph representations have been applied to cardiac electrophysiology for detecting reentrant circuits through directed activation analysis [Vandersickel et al., 2019], inferring causal relationships between fibrillatory signals [Handa, X. Li, et al., 2020], and spectral analysis of electrograms on patient-specific graphs [M. Sun et al., 2020], these approaches employ fixed graph structures and classical signal processing techniques rather than learning from data. While this thesis develops general-purpose graph-based spatiotemporal machine learning methods applicable beyond cardiology, the interplay between methodological development and the practical demands of AF mapping forms the central thread throughout.

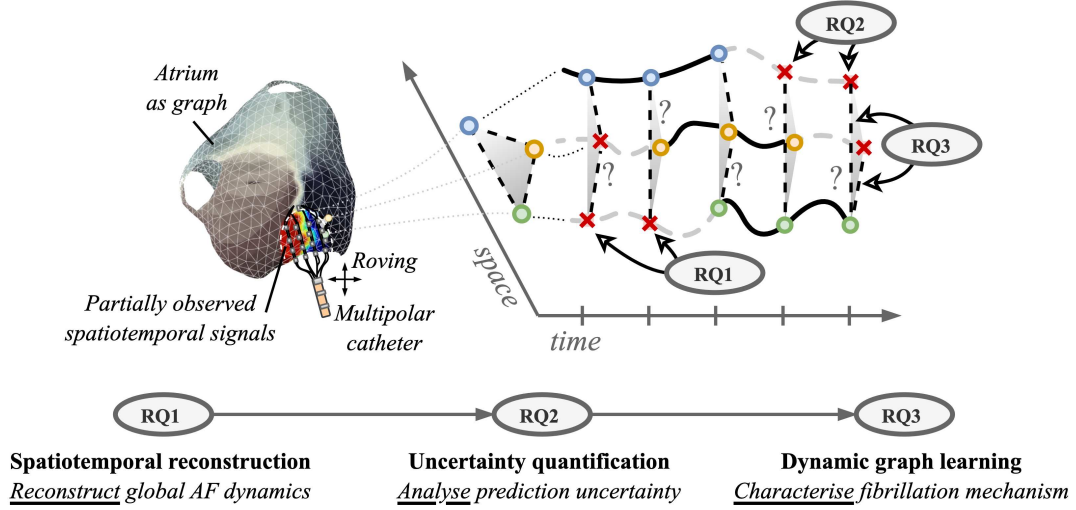


Figure 1.3: Atrial fibrillation mapping formulated as a graph-based spatiotemporal learning problem. Multipolar catheters provide sparse, sequential measurements of electrical activity across atrial tissue, where electrical potentials at observed locations form signals evolving over time with spatial relationships determined by the anatomical connectivity of the tissue (graph). Red crosses denote unobserved locations and time points, while dashed edges and question marks denote unknown or evolving functional connectivity. The three research questions (RQs) of this thesis build a progressive pipeline for AF mapping: **RQ1** *reconstructs* global AF dynamics from incomplete spatiotemporal observations; **RQ2** *analyses* prediction uncertainty through calibrated prediction intervals; and **RQ3** *characterises* fibrillation mechanisms by learning dynamic functional connectivity from streaming spatiotemporal signals.

1.2 Research Questions

This thesis aims to advance the methodological foundations of graph-based spatiotemporal machine learning and demonstrate their application to cardiac electrophysiology. The following research questions are addressed:

Research Question 1 (RQ1): *Can a spatiotemporal graph model reconstruct global atrial fibrillation dynamics from sparse catheter measurements while generalising across patients and adapting to individual cardiac substrates?*

Mapping AF requires visualising global electrical activation patterns across the atria, which could potentially inform catheter ablation therapy planning. Current machine learning approaches to AF mapping focus on classifying local electrogram features to guide ablation [Liao et al., 2021; Seitz et al., 2022; Manongi, J. Kim, and Goldbarg, 2023], but cannot provide a global view of wavefront dynamics across the atria. Methods that do provide whole-atrial coverage, such as non-contact inverse reconstruction [Narayan, Krummen, et al., 2012; Roney et al., 2017; Buch et al., 2016; Rudy and Burnes, 1999] and digital twin simulations [Boyle et al., 2019; Sakata et al., 2024] face bar-

riers to routine clinical use: the former suffers from the ill-posed nature of inverse problems and limited spatial resolution, while the latter requires pre-procedural imaging and extensive offline computation. These limitations motivate a fundamentally different approach: treating the reconstruction of global dynamics from sparse, sequential contact measurements as a spatiotemporal imputation problem, operating directly on data already collected during routine clinical procedures. Methodologically, this demands solutions to several interconnected challenges: (i) how to represent patient-specific atrial geometry as a graph amenable to learning, (ii) how to design models that are inductive over graph structure so that a single architecture generalises across patients with diverse anatomies, (iii) how to train reconstruction models when ground-truth global dynamics are unavailable during clinical procedures, and (iv) how to account for latent patient-specific properties such as tissue conductivity and fibrosis patterns that shape propagation but are not directly observed. This thesis provides a proof-of-concept for this imputation-based approach to mapping, leveraging graph-based spatiotemporal machine learning to enhance the visualisation of complex AF dynamics and facilitate the characterisation of fibrillation mechanisms from routine mapping data.

Research Question 2 (RQ2): *How can a relational graph structure between correlated time series be exploited to provide calibrated, distribution-free uncertainty quantification for spatiotemporal predictions?*

Beyond the accuracy of point predictions, the reliability of spatiotemporal forecasts is critical for effective decision-making in risk-sensitive applications [Makridakis, 1996; Petropoulos et al., 2022]. In the context of AF mapping, ablation guided by reconstructed activation patterns without quantified reliability risks treatment failure or patient harm. More broadly, existing probabilistic forecasting frameworks often rely on strong distributional assumptions and require major modifications to the base predictor [Benidis et al., 2022; Salinas et al., 2020], preventing their application to pre-trained models. Conformal prediction [Vovk, Gammerman, and Shafer, 2005; Angelopoulos, Bates, et al., 2023] offers a distribution-free alternative that can be applied on top of any pre-trained predictor without modification, providing finite-sample coverage guarantees from observed prediction residuals. Yet existing conformal methods for time series operate independently on each sequence and cannot account for the inter-series dependencies that characterise spatiotemporal data [Stankeviciute, M Alaa, and Schaar, 2021; C. Xu and Y. Xie, 2023a; C. Xu and Y. Xie, 2023b]. This is a missed opportunity: spatiotemporal dependencies offer both a challenge, as overlooking them can lead to wide prediction intervals, and an opportunity, as exploiting them may provide additional knowledge to reduce uncertainty [Zambon and Alippi, 2022]. Furthermore, the temporal non-stationarity inherent in many real-world systems means that the error

distribution may shift over time, requiring adaptive mechanisms to maintain valid coverage. This thesis introduces a conformal prediction framework that leverages graph deep learning operators to exploit relational structure between correlated time series, learning sparse dependency graphs directly from prediction residuals and incorporating adaptive components to handle non-stationary dynamics.

Research Question 3 (RQ3): *How can sparse, time-varying graph topologies be learned and tracked from streaming multivariate time series in non-stationary systems?*

Characterising the mechanisms sustaining cardiac fibrillation requires identifying preferential pathways of electrical propagation and assessing their temporal stability [Handa, X. Li, et al., 2020; F. S. Ng et al., 2020]. More broadly, many complex systems exhibit functional connectivity that evolves over time, from brain networks reconfiguring during cognitive tasks [Preti, Bolton, and Van De Ville, 2017] to financial networks whose causal structure shifts during market crises [Papana et al., 2017]. Existing approaches to graph topology estimation typically operate in batch mode [Friedman, Hastie, and Tibshirani, 2007; Xiaowen Dong, Thanou, et al., 2016; Mei and Moura, 2016], capturing only time-averaged connectivity patterns and requiring costly recomputation as new data arrives. In cardiac fibrillation, batch methods such as Granger causality identify few causal edges when dynamics are highly variable, as temporal variations are averaged out [Handa, X. Li, et al., 2020]. While deep learning approaches can learn graph structure end-to-end [Cini, Zambon, and Alippi, 2023; Shang and J. Chen, 2021], their computational demands and reliance on large training datasets make them unsuitable for real-time tracking of evolving topologies in streaming settings, where constant per-iteration complexity and immediate adaptation are essential. Even among existing online methods, achieving true sparsity remains a fundamental challenge: adaptive algorithms struggle to explicitly zero out non-causal connections despite employing sparsity-promoting regularisation [Taheri and Vorobyov, 2011; Zaman, Ramos, et al., 2020], yet the binary presence or absence of edges is precisely what determines graph topology and its causal interpretation. This thesis develops a sparsity-aware adaptive algorithm that learns time-varying graph topologies from streaming multivariate time series with constant per-iteration cost, enabling real-time tracking of functional connectivity in non-stationary systems with application to characterising the structure and stability of cardiac fibrillation.

1.3 Main Contributions

This thesis is based on several main contributions with the overarching goal to address the research questions articulated above. The thesis advances both the methodological foundations of

graph-based spatiotemporal machine learning and demonstrates their application to cardiac electrophysiology through four interconnected studies.

PULSEDIFF: Improving Diffusion Models for ECG Imputation (Chapter 4). PULSEDIFF is a template-guided denoising diffusion probabilistic model for imputing missing values in electrocardiogram (ECG) signals, introduced as an exploration of how temporal priors can guide state-of-the-art generative models for physiological signal imputation. While diffusion models have shown promise for time series imputation, existing approaches do not exploit the quasi-periodic structure inherent in physiological signals or account for subject-specific variations. PULSEDIFF addresses these limitations through three key steps: (1) extraction of subject-level waveform templates from observed data to provide informative, personalised temporal priors; (2) beat-level stochastic augmentation to capture variations in position and amplitude across cardiac cycles; and (3) a confidence scoring mechanism that adapts the prior strength based on signal regularity and health condition. Validated on the PTB-XL dataset [Wagner et al., 2020], PULSEDIFF improves upon strong baselines for diffusion models in time series and outperforms leading deterministic methods for short-interval missing data while remaining competitive for long-interval gaps. The chapter further establishes that such temporal priors are insufficient when quasi-periodicity assumptions break down, motivating the incorporation of spatial structure as a relational inductive bias in subsequent chapters, where graph-based representations of tissue connectivity enable reconstruction even when temporal patterns are irregular.

FIBMAP: Learning to Reconstruct Global Atrial Fibrillation Dynamics From Sparse Measurements (Chapter 5, RQ1). FIBMAP is a graph recurrent neural network that reconstructs global atrial fibrillation dynamics from sparse, sequential catheter measurements, formulating cardiac mapping as a spatiotemporal graph signal reconstruction problem where electrical potentials at tissue locations form signals on a graph determined by anatomical connectivity. A training strategy that simulates realistic mapping procedures by extracting sparse subsets from high-resolution ground-truth data enables the model to learn reconstruction from limited observations. Validated on high-resolution *ex vivo* optical mapping of rat ventricular fibrillation (17 hearts, 253 million observations) and clinical non-contact mapping of human AF (51 patients, 4 billion observations), FIBMAP reconstructs global dynamics from only 10% instantaneous coverage, outperforming baselines by 47–53% in mean absolute error and 23–109% in phase singularity detection. Personalised reconstruction is further demonstrated from real-world routine contact mapping data, showcasing clinical feasibility. This chapter directly addresses **RQ1** by establishing graph-based spatiotemporal frameworks for patient-specific AF mapping and demonstrating their clinical viability.

COREL: Relational Conformal Prediction for Calibrated Uncertainty Quantification (Chapter 6, RQ2). COREL is a distribution-free approach for uncertainty quantification in spatiotemporal forecasting that exploits relational structure between correlated time series, training a spatiotemporal graph neural network on prediction residuals from a calibration set to approximate the quantile function of the error distribution. Uncertainty estimates are conditioned on observations at neighbouring nodes as defined by a learned sparse dependency graph, requiring no prior knowledge of relational structure and permitting application on top of any pre-trained predictor without modification. An adaptive component further updates local node embeddings over time to handle non-stationary error distributions. Applied to traffic, air quality, and energy consumption benchmarks, COREL achieves state-of-the-art uncertainty quantification with accurate coverage, directly addressing **RQ2**. COREL was validated on general forecasting benchmarks rather than cardiac data. Extending the method from the forecasting to the imputation setting and integrating it with FIBMAP to provide calibrated prediction intervals for imputation maps is an identified direction for future work (Chapter 8).

ADACGP: Online Graph Topology Learning via Time-Vertex Adaptive Filters (Chapter 7, RQ3). ADACGP is a sparsity-aware adaptive algorithm for learning time-varying graph topologies from multivariate time series. The graph shift operator is estimated through recursive update formulae that addresses sparsity, shift-invariance, and estimation bias, enabling real-time tracking of connectivity changes with constant per-iteration computational cost. A key innovation is a variable splitting approach that achieves true sparsity in the online setting, reliably identifying the presence or absence of causal connections where existing adaptive methods cannot. Through comprehensive simulations across diverse graph topologies, ADACGP achieves improvements exceeding 83% in topology estimation compared to existing methods while maintaining favourable computational scaling. Applied to *ex vivo* optical mapping recordings of rat ventricular fibrillation, ADACGP tracks dynamic propagation patterns more effectively than Granger causality, identifying temporal variations in functional connectivity that static approaches miss, and characterises the stability of conduction patterns that may sustain the arrhythmia. This chapter addresses **RQ3** by enabling the discovery and tracking of functional connectivity in non-stationary spatiotemporal systems, with the fibrillation application demonstrating that dynamic connectivity characterisation from cardiac recordings is achievable where existing batch methods fail.

Together, these contributions establish a comprehensive framework for graph-based spatiotemporal machine learning: reconstructing signals from sparse observations (FIBMAP), quantifying prediction uncertainty (COREL), and learning dynamic functional connectivity (ADACGP). FIBMAP represents a task-specific architecture tailored to the unique challenges of AF mapping,

demonstrating the viability of graph-based spatiotemporal approaches in this domain. In contrast, COREL and ADACGP are general-purpose methodologies applicable across diverse spatiotemporal problems beyond cardiac electrophysiology. While cardiac applications motivate the methods throughout this thesis, the underlying principles and algorithms extend to any domain requiring the analysis of time-varying signals on irregular network structures.

1.4 Thesis Organisation

The thesis is organised into three parts, progressing from foundational background through signal reconstruction to analysis and characterisation of spatiotemporal systems. Figure 1.4 provides a visual overview of this structure and the relationships between chapters.

Part I — Background establishes the theoretical and clinical foundations. **Chapter 2** covers discrete-time signal processing, graph signal processing, and graph-based spatiotemporal machine learning, providing the mathematical framework for the methods developed in this thesis. **Chapter 3** establishes the clinical and electrophysiological foundations for the thesis. It covers cardiac electrical activity from the cellular to body-surface scale, tissue-level propagation dynamics, the mechanisms and electrophenotype spectrum of atrial fibrillation, and current treatment strategies including catheter ablation. The chapter concludes with a survey of electrophysiological mapping modalities, highlighting the fundamental resolution–coverage trade-off that motivates the data-driven reconstruction methods developed in subsequent chapters.

Part II — Imputation and Reconstruction addresses the challenge of recovering signals from sparse, incomplete observations. **Chapter 4** explores temporal priors for physiological signal imputation through PULSEDIFF, a template-guided diffusion model for ECG reconstruction. Waveform templates serving as quasi-periodic temporal priors substantially improve imputation quality, while the breakdown of such regularity motivates the shift to relational inductive biases through graph-based spatial representations in subsequent chapters. **Chapter 5** presents FIBMAP, the central contribution of this thesis, which formulates cardiac mapping as spatiotemporal graph signal reconstruction and develops methods to recover global atrial fibrillation dynamics from sparse catheter measurements, with validation on optical mapping data and clinical recordings.

Part III — Analysis and Characterisation of Spatiotemporal Systems develops general-purpose methodologies for analysing and characterising time-varying signals on graphs. **Chapter 6** introduces COREL, a distribution-free conformal prediction framework that exploits relational graph structure among correlated time series to provide calibrated, adaptive uncertainty quantification for spatiotemporal forecasts without modification of the base predictor. **Chapter 7** develops

ADACGP, a sparsity-aware adaptive algorithm for learning time-varying graph topologies from streaming multivariate time series, enabling real-time tracking of functional connectivity with application to characterising dynamic propagation patterns in cardiac fibrillation.

Finally, **Chapter 8** concludes with a discussion of broader implications and future directions, including potential extensions such as integrating uncertainty quantification with reconstruction and leveraging graph partitioning techniques to simulate ablation procedures for treatment planning.

1.5 List of Publications

This thesis is based on research initially published by the author in a set of peer-reviewed publications and manuscripts under review. In the following, each publication and its original appearance is listed, and the author’s contribution to the respective works is outlined in terms of conceptualisation, experimental design and implementation, and writing.

- **Alexander Jenkins***, Zehua Chen*, Fu Siong Ng and Danilo Mandic (2023). “Improving Diffusion Models for ECG Imputation with an Augmented Template Prior.” In: *ICML Workshop on Interpretable Machine Learning in Healthcare* (*equal contribution), *cf.* Chapter 4.

Contribution: The author conceptualised the solution alongside ZC, designed and performed the experiments, and wrote the manuscript.

- **Alexander Jenkins**, Andrea Cini, Joseph Barker, Alexander Sharp, Arunahis Sau, Varun Valentine, Srushti Valasang, Xinyang Li, Tom Wong, Timothy Betts, Danilo Mandic, Cesare Alippi and Fu Siong Ng (2026). “Learning to Map Global Atrial Fibrillation Dynamics from Sparse Measurements” *EP Europace* (**under review**), *cf.* Chapter 5.

Contribution: The author conceptualised the solution alongside AC, designed and performed the experiments, and wrote the manuscript.

- **Alexander Jenkins**, Andrea Cini, Joseph Barker, Alexander Sharp, Arunahis Sau, Varun Valentine, Srushti Valasang, Xinyang Li, Tom Wong, Timothy Betts, Danilo Mandic, Cesare Alippi and Fu Siong Ng (2026). “Mapping Global Dynamics of Atrial Fibrillation from Sparse Measurements with Spatiotemporal Graph Neural Networks.” In: *Heart Rhythm 2026*, *cf.* Chapter 5.

Contribution: The author conceptualised the solution alongside AC, designed and performed the experiments, and wrote the manuscript.

- **Alexander Jenkins**, Andrea Cini, Joseph Barker, Danilo Mandic, Cesare Alippi and Fu Siong Ng (2025). “Mapping the Dynamics of Atrial Fibrillation with Spatiotemporal Graph Neural Networks.” In: *NeurIPS Workshop on Learning from Time Series for Health*, cf. Chapter 5.

Contribution: The author conceptualised the solution alongside AC, designed and performed the experiments, and wrote the manuscript.

- Andrea Cini, **Alexander Jenkins**, Danilo Mandic, Cesare Alippi and Filippo Maria Bianchi (2025). “Relational Conformal Prediction for Correlated Time Series.” In: *International Conference on Machine Learning (ICML)*, cf. Chapter 6.

Contribution: The author contributed to the conceptualisation of the solution, designed and performed experiments, and assisted in writing and editing the manuscript.

- **Alexander Jenkins**, Thiernithi Varidddhisai, Ahmed El-Medany, Fu Siong Ng and Danilo Mandic (2025). “Online Graph Topology Learning via Time-Vertex Adaptive Filters: From Theory to Cardiac Fibrillation.” In: *IEEE Transactions on Signal and Information Processing over Networks*, vol. 11, pp. 965–979, cf. Chapter 7.

Contribution: The author conceptualised the solution alongside TV, designed and performed the experiments, and wrote the manuscript.

The methods developed in this thesis have direct clinical translational relevance, and the author has filed patent applications to protect the intellectual property associated with the imputation approach to cardiac mapping. These applications cover the methodology for reconstructing global cardiac dynamics from sparse measurements:

- **Alexander Jenkins**, Andrea Cini, Danilo Mandic, Cesare Alippi and Fu Siong Ng. “A Method of Constructing Maps of Dynamical Variables on a Cardiac Surface.” UK Patent Application No. GB2500830.1, filed January 2025, cf. Chapter 5.
- **Alexander Jenkins**, Andrea Cini, Danilo Mandic, Cesare Alippi and Fu Siong Ng. “A Method of Constructing Maps of Dynamical Variables on a Cardiac Surface.” PCT International Patent Application No. PCT/GB2026/050073, filed January 2026, cf. Chapter 5.

In addition, the author has contributed to the following publications. These describe methods and applications of graph-based methods to other domains or collaborative clinical work outside the main scope of this thesis.

- **Alexander Jenkins**, Emma Rainbow, Imad Jaimoukha, Fu Siong Ng, Ljubisa Stankovic,

Milos Dakovic, Danimir Mandic and Danilo Mandic (2026). “A Graph-Theoretic Approach to Fair and Skill-Diverse Student Group Formation.” In: *IEEE Signal Processing Magazine*, January 2026.

Contribution: The author conceptualised the solution, designed and performed the experiments, and wrote the manuscript.

- **Alexander Jenkins**, Imad Jaimoukha, Ljubisa Stankovic and Danilo Mandic (2023). “Fair and Skill-Diverse Student Group Formation via Constrained k -Way Graph Partitioning.” In: *AAAI Workshop on AI for Education (contributed talk)*.

Contribution: The author conceptualised the solution, designed and performed the experiments, and wrote the manuscript.

- Joseph Barker, Arunashis Sau, Nikesh Bajaj, **Alexander Jenkins**, Alex Sharp, Xili Shi, Xinyang Li, Nabeela Karim, Balvinder Handa, Richard Chambers, Timothy Betts, Nicholas S. Peters, Tom Wong and Fu Siong Ng (2025). “Granger Causality Connectivity Analysis of Persistent Atrial Fibrillation Dynamics Reveals Posterior Wall Mechanistic Insights.” In: *Heart Rhythm O2*, vol. 6, no. 8, pp. 1097–1105.

Contribution: The author assisted in the writing and editing of the manuscript.

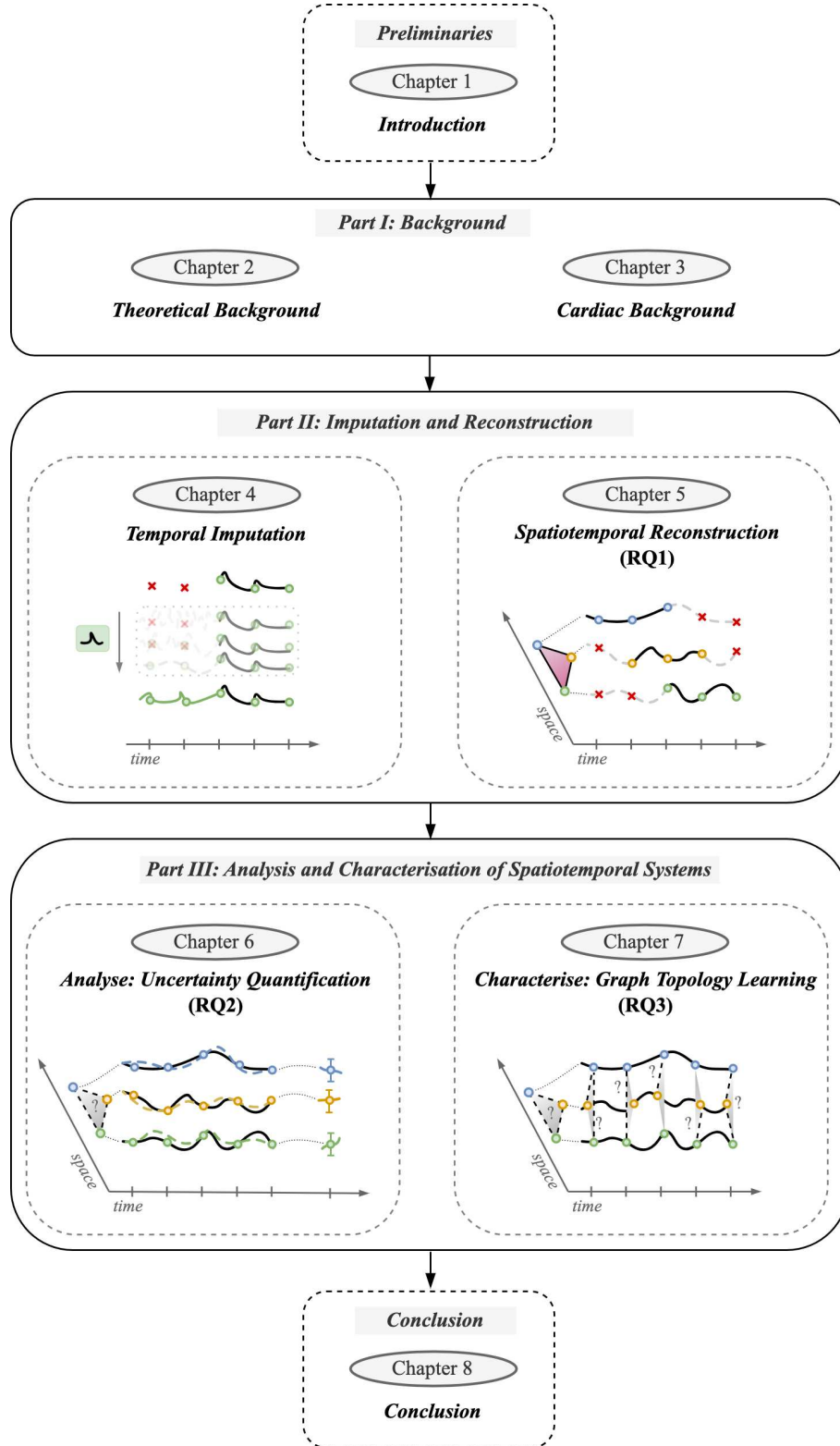


Figure 1.4: Thesis structure and conceptual overview. The thesis is organised into three parts: background (Part I), imputation and reconstruction (Part II), and analysis and characterisation of spatiotemporal systems (Part III). The three main contributions form a progressive pipeline: *reconstruct* global dynamics from sparse measurements (Chapter 5, **RQ1**), *analyse* prediction reliability through calibrated uncertainty quantification (Chapter 6, **RQ2**), and *characterise* fibrillation mechanisms by learning dynamic functional connectivity (Chapter 7, **RQ3**). Temporal imputation (Chapter 4) demonstrates the promise and limitations of temporal inductive biases for physiological signal reconstruction, motivating the shift to relational inductive biases through graph-based spatial representations in subsequent chapters.

Part I

Background

Chapter 2

Theoretical Background

This chapter establishes the theoretical foundations for the graph-based spatiotemporal machine learning methods developed in this thesis. The chapter is organised in three parts that progressively generalise from regular to irregular domains. First, discrete-time signal processing introduces the shift operator, linear time-invariant systems, autoregressive models, adaptive filtering, and neural sequence models on regular temporal grids. Second, graph signal processing extends these concepts to irregular network structures through the graph shift operator, polynomial graph filters, spectral graph theory, and graph neural networks. Third, graph-based spatiotemporal models combine temporal and graph-domain operations to jointly process signals evolving over both time and irregular spatial topologies, from classical graph autoregressive processes to modern spatiotemporal graph neural network architectures. The chapter concludes by identifying three shortfalls of the standard framework, namely the assumption of homogeneous node dynamics, known graph topologies, and deterministic point predictions, that motivate the methods developed in the results chapters of this thesis.

2.1	Signals and Systems on Regular Domains: Time Series Models	17
2.2	Signals and Systems on Irregular Domains: Graph Signal Processing	27
2.3	Time-Varying Signals and Systems on Irregular Domains: Graph-Based Spatiotemporal Models	39

2.1 Signals and Systems on Regular Domains: Time Series Models

Continuous-time signals are ubiquitous in the physical world, describing phenomena from electromagnetic wave propagation to atmospheric dynamics. While these analogue signals are the natural representation of physical processes, they are routinely sampled and converted to discrete-time sequences for digital signal processing, enabling robust transmission, exact reproducibility, and flexible algorithmic implementation [Oppenheim, Schafer, and Buck, 1999].

A discrete-time signal, $\{x[n]\}_{n \in \mathbb{Z}}$, assigns a value to each integer time index (see Fig. 2.1), either

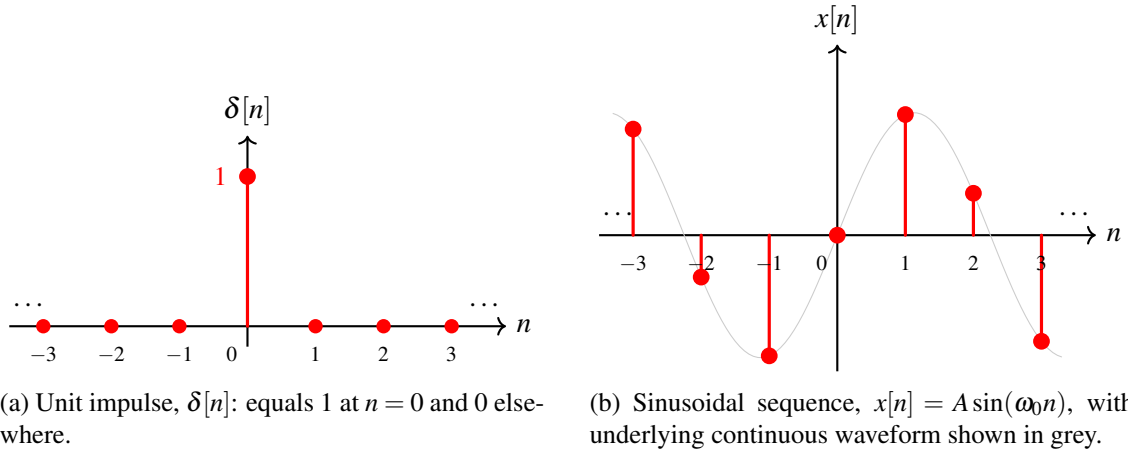


Figure 2.1: Illustration of discrete-time signals. A discrete-time signal assigns a value to each integer time index n . (a) The unit impulse is a fundamental building block for representation of signals and systems in digital signal processing (DSP). (b) A sinusoidal sequence obtained by sampling a continuous-time sinusoid, illustrating how discrete samples capture the behaviour of an underlying analogue signal.

arising from an inherently discrete process or obtained by uniformly sampling a continuous-time signal, $x_c(t)$, at interval Δt , such that $x[n] = x_c(n\Delta t)$. The Nyquist-Shannon sampling theorem guarantees that perfect reconstruction of $x_c(t)$ from its samples is possible provided the sampling frequency, $f_s = 1/\Delta t$, satisfies $f_s \geq 2f_{\max}$, where $f_{\max}$ is the highest frequency present in the signal [Oppenheim, Schaffer, and Buck, 1999].

This section outlines the core concepts of discrete-time signal processing that underpin the graph-based spatiotemporal methods of this thesis. The focus is on three ideas that recur throughout later chapters: the shift operator and its role in defining convolution, autoregressive models as parametric signal representations, and neural sequence models that extend linear methods to non-stationary, non-linear settings. These temporal concepts provide the direct foundation for their generalisation to irregular graph-structured domains in [Sec. 2.2](#).

2.1.1 Linear Time-Invariant Systems and the Shift Operator

A discrete-time system, $\mathcal{T}$, maps an input sequence, $x[n]$, to an output sequence, $y[n] = \mathcal{T}\{x[n]\}$. The most analytically tractable class comprises systems that are both linear and time-invariant. Linearity requires that the system satisfies the principle of superposition, which states that for any signals $x_1[n], x_2[n]$ and scalars a_1, a_2 , the following holds

$$\mathcal{T}\{a_1 x_1[n] + a_2 x_2[n]\} = a_1 \mathcal{T}\{x_1[n]\} + a_2 \mathcal{T}\{x_2[n]\}. \quad (2.1)$$

This combines additivity (responses to separate inputs sum) and homogeneity (scaling the input scales the output proportionally). Time invariance requires that a time shift in the input produces an identical shift in the output. Concretely, if $y[n] = \mathcal{T}\{x[n]\}$, then

$$\mathcal{T}\{x[n - n_0]\} = y[n - n_0], \quad (2.2)$$

for all integer delays, n_0 . This reflects systems whose input-output characteristics remain constant over time, in contrast to adaptive or time-varying systems whose parameters may evolve.

Systems satisfying both properties, known as linear time-invariant (LTI) systems, are completely characterised by their impulse response, $h[n] = \mathcal{T}\{\delta[n]\}$, where $\delta[n]$ denotes the unit impulse [Oppenheim, Schafer, and Buck, 1999]. Since any signal can be decomposed as a weighted sum of shifted impulses, applying linearity and time invariance yields the convolution sum,

$$y[n] = \sum_{k=-\infty}^{+\infty} x[k] h[n - k] = x[n] * h[n], \quad (2.3)$$

where $*$ denotes the convolution operator. Convolution is commutative, associative, and distributive over addition, with the unit impulse $\delta[n]$ acting as the identity element. These properties enable cascaded LTI systems to be analysed as a single equivalent system and parallel systems to be combined through summation.

For LTI systems, causality and stability can be expressed directly in terms of the impulse response. Such a system is causal if and only if $h[n] = 0$ for all $n < 0$, ensuring the output does not depend on future inputs, and bounded-input bounded-output stable if and only if its impulse response is absolutely summable, $\sum_n |h[n]| < \infty$.

The Time-Shift Operator

The time delay is a fundamental operation in DSP. The ideal delay system shifts a signal by n_0 samples, $y[n] = x[n - n_0]$. Its impulse response is $h[n] = \delta[n - n_0]$, demonstrating that delay operations are themselves LTI systems. This observation motivates the introduction of the time-shift operator, z^{-1} , illustrated in Fig. 2.2, defined by

$$(z^{-1}x)[n] = x[n - 1], \quad (2.4)$$

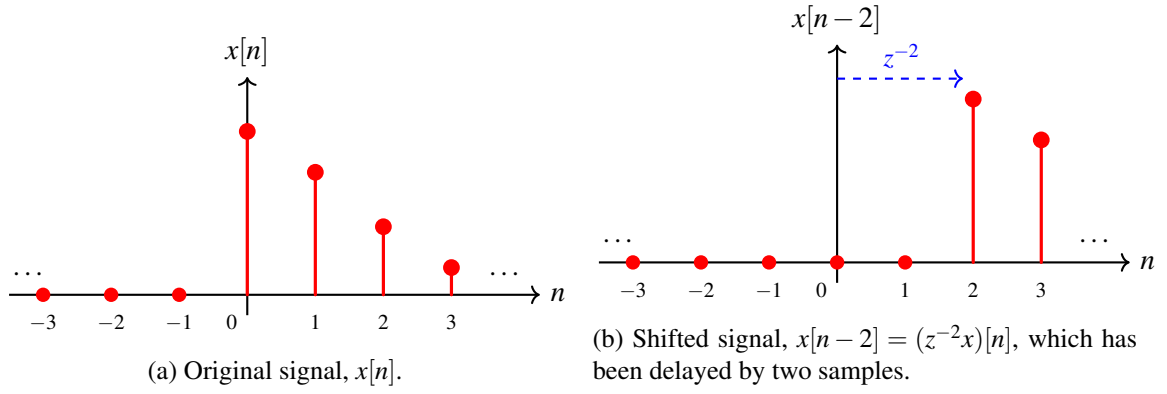


Figure 2.2: Illustration of the time-shift operator. (a) An arbitrary causal signal $x[n]$. (b) The same signal after applying the shift operator twice, z^{-2} , delaying each sample by two time steps. The dashed arrow indicates the direction of the shift.

where repeated application yields k -step delays, $(z^{-k}x)[n] = x[n-k]$. Using this notation, a finite impulse response (FIR) filter of order M can be written as a polynomial in z^{-1} ,

$$y[n] = \sum_{k=0}^M h[k]x[n-k] = h(z^{-1})x[n], \quad \text{where} \quad h(z^{-1}) = \sum_{k=0}^M h[k]z^{-k}. \quad (2.5)$$

FIR filters are always stable and have finite temporal support. Infinite impulse response (IIR) filters extend this with feedback (recursive) terms, yielding rational functions of z^{-1} ,

$$A(z^{-1})y[n] = B(z^{-1})x[n], \quad (2.6)$$

where $A(z^{-1}) = 1 + \sum_{k=1}^P a_k z^{-k}$ and $B(z^{-1}) = \sum_{k=0}^Q b_k z^{-k}$. Stability of IIR filters requires that the roots of $A(z^{-1})$ lie inside the unit circle [Oppenheim, Schaffer, and Buck, 1999].

Remark 1. *The polynomial representation reveals a key structural insight. Discrete-time filtering is algebraically equivalent to applying polynomials in the shift operator to the input signal. This perspective connects directly to graph signal processing (Sec. 2.2.3), where the temporal shift operator, z^{-1} , is replaced by a graph shift operator, $\mathbf{S}$, and graph filters are similarly expressed as polynomials in $\mathbf{S}$.*

Frequency Response

Complex exponentials, $e^{j\omega n}$, are eigenfunctions of the time-shift operator, since $z^{-1}[e^{j\omega n}] = e^{-j\omega} e^{j\omega n}$. Because LTI systems are polynomials (or rational functions) of z^{-1} , complex exponentials are also eigenfunctions of any LTI system, yielding $y[n] = H(e^{j\omega})e^{j\omega n}$ when $x[n] = e^{j\omega n}$. The scaling factor,

$$H(e^{j\omega}) = \sum_{k=-\infty}^{+\infty} h[k]e^{-j\omega k} \quad (2.7)$$

is known as the frequency response, which is the discrete-time Fourier transform of the impulse response. Convolution in the time domain corresponds to pointwise multiplication in the frequency domain: $Y(e^{j\omega}) = H(e^{j\omega})X(e^{j\omega})$. Filter design thus reduces to specifying the frequency response, $H(e^{j\omega})$, to selectively amplify or attenuate frequency components.

Remark 2. *This operator-centric view highlights that the algebraic role of frequency in DSP (diagonalising LTI systems and reducing convolution to pointwise multiplication) arises from the eigenstructure of the time shift operator rather than from the temporal domain itself. On regular temporal grids, the shift is uniform and its eigenfunctions are known analytically as complex exponentials. On irregular graph domains, however, there is no canonical shift, and the eigenfunctions must instead be derived from the graph topology through eigendecomposition of the graph shift operator. This motivates the graph Fourier transform (GFT) developed in [Sec. 2.2.4](#).*

2.1.2 Random Signals and Stationarity

Many practical signals exhibit inherent randomness and are modelled as discrete-time stochastic processes, $\{x[n]\}$, where each $x[n]$ is a random variable. A stochastic process is wide-sense stationary (WSS) if its mean is constant, $\mathbb{E}[x[n]] = m_x$ for all n , and its autocorrelation depends only on the time lag,

$$\phi_{xx}[k] = \mathbb{E}[x[n]x^*[n-k]], \quad \forall n, \quad (2.8)$$

where $*$ denotes complex conjugation. WSS, combined with the assumption of ergodicity (time averages converge to ensemble averages), permits statistical estimation from a single observed realisation, which is essential in practice. A fundamental example of a stochastic process is white noise, a zero-mean, uncorrelated process $\{e[n]\}$ with $\phi_{ee}[k] = \sigma^2\delta[k]$, containing equal power at all frequencies. White noise serves as the innovation process driving autoregressive models and represents the irreducible prediction error in optimal filtering.

Remark 3. *The stationarity assumption underpins much of classical signal processing but is violated by many real-world signals whose statistics evolve over time. The adaptive and neural methods discussed in [Sections 2.1.4 and 2.1.5](#) address this limitation. Its graph-domain analogue, time-vertex stationarity, is introduced in [Sec. 2.3.1](#).*

2.1.3 Autoregressive Models

Autoregressive (AR) models provide a parametric framework for representing stochastic processes through recursive difference equations. An $\text{AR}(P)$ model expresses the current sample as a linear

combination of P past values plus a white noise innovation,

$$x[n] = \sum_{k=1}^P a_k x[n-k] + e[n] = \mathbf{a}^\top \mathbf{x}_P[n] + e[n], \quad (2.9)$$

where $\mathbf{a} = [a_1, \dots, a_P]^\top$ collects the AR coefficients and $\mathbf{x}_P[n] = [x[n-1], \dots, x[n-P]]^\top$ collects the lagged values. From the perspective of [Sec. 2.1.1](#), AR models are IIR filters driven by white noise, with the feedback structure of (2.6) and $B(z^{-1}) = 1$. The optimal AR coefficients under the mean-square error criterion, $\mathbf{a}_o$, satisfy the Yule–Walker equations [[Hayes, 1996](#)],

$$\mathbf{R} \mathbf{a}_o = \mathbf{r}, \quad (2.10)$$

where $\mathbf{R}$ is the $P \times P$ autocorrelation matrix with entries $R_{kl} = \phi_{xx}[k-l]$ for $k, l \in \{1, \dots, P\}$, and $\mathbf{r} = [\phi_{xx}[1], \dots, \phi_{xx}[P]]^\top$ is the autocorrelation vector.

More generally, the autoregressive moving average model, $\text{ARMA}(P, Q)$, combines AR feedback with moving average (MA) forward terms,

$$x[n] = \sum_{k=1}^P a_k x[n-k] + \sum_{j=1}^Q b_j e[n-j] + e[n], \quad (2.11)$$

or in operator form, $A(z^{-1})x[n] = B(z^{-1})e[n]$. Pure MA models ($P = 0$) are FIR filters; ARMA models combine the parsimony of AR representation with the flexibility of MA terms.

Vector Autoregressive Models

The vector autoregressive (VAR) model extends (2.9) to jointly model N coupled time series, $\mathbf{x}[n] \in \mathbb{R}^N$ (see [Fig. 2.3](#)),

$$\mathbf{x}[n] = \sum_{k=1}^P \mathbf{A}_k \mathbf{x}[n-k] + \mathbf{e}[n], \quad (2.12)$$

where $\mathbf{A}_k \in \mathbb{R}^{N \times N}$ are coefficient matrices whose off-diagonal elements capture directed cross-channel relationships from series j to series i at lag k . Non-zero off-diagonal entries form the basis for Granger causality analysis [[Granger, 1969](#)], quantifying whether the history of one series improves prediction of another beyond its own past. In operator form, $\mathbf{A}(z^{-1})\mathbf{x}[n] = \mathbf{e}[n]$ where $\mathbf{A}(z^{-1}) = \mathbf{I} - \sum_{k=1}^P \mathbf{A}_k z^{-k}$.

Remark 4. *The VAR formulation directly motivates graph-based spatiotemporal models. When the N time series reside on a graph structure, the dense coefficient matrices $\mathbf{A}_k$ can be replaced by polynomial graph filters of order K ([Sec. 2.3.1](#)), reducing parameters from $\mathcal{O}(PN^2)$ to $\mathcal{O}(PK)$ while encoding the spatial structure of the underlying network. This connection is developed in*

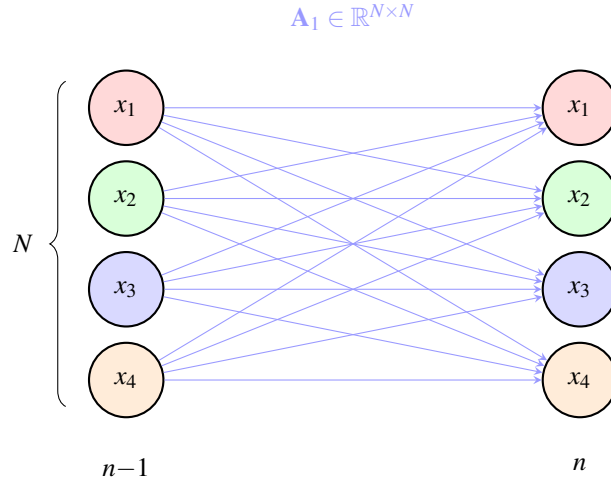
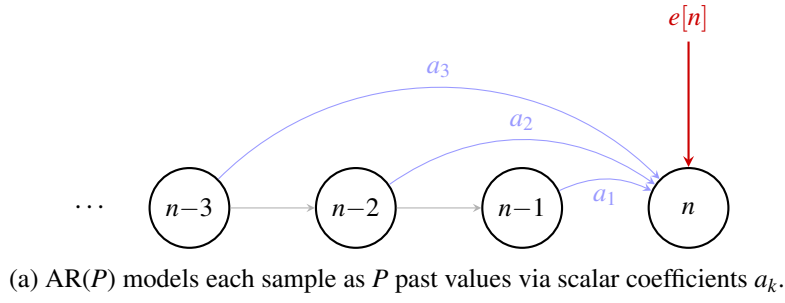


Figure 2.3: Scalar and vector autoregressive models. (a) A scalar AR(P) process expresses the current sample as a weighted sum of P past values. (b) The VAR model extends this to N coupled series with dense coefficient matrices $\mathbf{A}_k$, where every series may influence every other.

Sec. 2.3.

2.1.4 Adaptive Filtering

The Yule–Walker solution (2.10) assumes known second-order statistics, which must be estimated in practice and may change over time. Adaptive filters address both limitations by updating their coefficients recursively online as new data arrives in a streaming setting, enabling operation on signals with unknown or non-stationary statistics [Haykin, 2008].

Consider the system identification setting, $d[n] = \mathbf{x}_P[n]^\top \mathbf{w}_o + v[n]$, where $d[n]$ is the desired signal, $\mathbf{x}_P[n] = [x[n-1], \dots, x[n-P]]^\top$ is the input regressor, $\mathbf{w}_o$ is the true (unknown) weight vector, and $v[n]$ is observation noise. An adaptive filter produces the estimate, $\hat{d}[n] = \mathbf{x}_P[n]^\top \mathbf{w}[n]$, and observes the output error, $e[n] = d[n] - \hat{d}[n]$. The weight vector that minimises the mean squared

error (MSE), $\xi = \mathbb{E}\{|e[n]|^2\}$, satisfies the Wiener–Hopf equation [Widrow and Stearns, 1985],

$$\mathbf{R} \mathbf{w}_o = \mathbf{p}, \quad (2.13)$$

where $\mathbf{R} = \mathbb{E}\{\mathbf{x}_P[n]\mathbf{x}_P[n]^\top\}$ is the input autocorrelation matrix, and $\mathbf{p} = \mathbb{E}\{d[n]\mathbf{x}_P[n]\}$ is the cross-correlation vector between the desired signal and the input. This generalises the Yule–Walker system (2.10), which arises as a special case when the desired signal is the process itself, $d[n] = x[n]$. The batch solution, $\mathbf{w}_o = \mathbf{R}^{-1}\mathbf{p}$, requires explicit knowledge of these statistics, motivating iterative alternatives.

Least Mean Square

The least mean square (LMS) algorithm [Widrow and Stearns, 1985] approximates the true gradient, $\nabla_{\mathbf{w}}\xi = -2(\mathbf{p} - \mathbf{R}\mathbf{w}[n])$, with an instantaneous stochastic estimate, yielding the weight update equation

$$\mathbf{w}[n+1] = \mathbf{w}[n] + \mu e[n]\mathbf{x}_P[n], \quad (2.14)$$

where $\mu > 0$ is the step size. Under stationarity, the LMS converges to the Wiener solution, $\mathbf{w}_o$, with convergence in the mean guaranteed for $0 < \mu < 2/\lambda_{\max}$, where $\lambda_{\max}$ is the largest eigenvalue of $\mathbf{R}$ [Haykin, 2008]. The convergence speed is governed by the eigenvalue spread of $\mathbf{R}$; the algorithm converges quickly for white inputs (equal eigenvalues) and slowly for coloured inputs. Its computational cost is $\mathcal{O}(P)$ per sample.

Recursive Least Squares

The recursive least squares (RLS) algorithm [Haykin, 2008] minimises a weighted least squares cost function with exponential forgetting,

$$C(\mathbf{w}[n]) = \sum_{\tau=1}^n \lambda^{n-\tau} |e[\tau]|^2, \quad (2.15)$$

where $\lambda \in (0, 1]$ is the forgetting factor that exponentially down-weights past observations. The minimiser satisfies $\mathbf{R}[n]\mathbf{w}[n] = \mathbf{p}[n]$, where

$$\mathbf{R}[n] = \sum_{\tau=1}^n \lambda^{n-\tau} \mathbf{x}_P[\tau]\mathbf{x}_P[\tau]^\top, \quad \mathbf{p}[n] = \sum_{\tau=1}^n \lambda^{n-\tau} d[\tau]\mathbf{x}_P[\tau], \quad (2.16)$$

are the exponentially weighted autocorrelation matrix and cross-correlation vector, respectively, with the weighting, $\lambda^{n-\tau}$, assigning exponentially decreasing influence to observations further in

the past. Both admit recursive updates,

$$\mathbf{R}[n] = \lambda \mathbf{R}[n-1] + \mathbf{x}_P[n] \mathbf{x}_P[n]^\top, \quad \mathbf{p}[n] = \lambda \mathbf{p}[n-1] + d[n] \mathbf{x}_P[n], \quad (2.17)$$

and applying the matrix inversion lemma to (2.17) gives a recursive update for the inverse,

$$\mathbf{R}[n]^{-1} = \lambda^{-1} \left(\mathbf{R}[n-1]^{-1} - \frac{\mathbf{R}[n-1]^{-1} \mathbf{x}_P[n] \mathbf{x}_P[n]^\top \mathbf{R}[n-1]^{-1}}{\lambda + \mathbf{x}_P[n]^\top \mathbf{R}[n-1]^{-1} \mathbf{x}_P[n]} \right), \quad (2.18)$$

avoiding explicit matrix inversion at each step. Defining the gain vector as $\mathbf{g}[n] = \mathbf{R}[n-1]^{-1} \mathbf{x}_P[n] / (\lambda + \mathbf{x}_P[n]^\top \mathbf{R}[n-1]^{-1} \mathbf{x}_P[n])$, the weight update takes the form

$$\mathbf{w}[n] = \mathbf{w}[n-1] + \mathbf{g}[n] \alpha[n], \quad (2.19)$$

where $\alpha[n] = d[n] - \mathbf{x}_P[n]^\top \mathbf{w}[n-1]$ is the a priori error. The RLS achieves convergence rates that are largely independent of the eigenvalue spread of $\mathbf{R}$, at the cost of $\mathcal{O}(P^2)$ complexity per sample [Haykin, 2008].

Remark 5. *The forgetting factor, λ , controls the effective memory of the filter: smaller values increase sensitivity to recent data, enabling tracking of time-varying systems at the expense of higher estimation variance. Setting $\lambda = 1$ recovers the growing-window least squares solution, equivalent to the batch Wiener–Hopf result as $n \rightarrow \infty$. As shown by Danilo P Mandic, Kanna, and Constantinides (2015), the LMS and RLS algorithms are intimately related to the Kalman filter. The extension of adaptive filtering to graph-structured data is central to Chapter 7.*

2.1.5 Neural Sequence Models

While AR and ARMA models provide optimal linear predictors, many real-world signals exhibit non-linear dynamics that linear combinations of past observations cannot capture. Neural sequence models address this by introducing learnable non-linear transformations.

Non-linear Autoregressive Models

The non-linear autoregressive model generalises (2.9) by replacing the linear combination with a non-linear function,

$$x[n] = f_\theta(x[n-1], \dots, x[n-P]) + e[n], \quad (2.20)$$

where $f_\theta(\cdot)$ may be realised by a feed-forward neural network, enabling approximation of arbitrary non-linear dependencies [S. Chen, Billings, and Grant, 1990]. However, explicitly specifying the lag order, P , remains a limitation.

Recurrent Neural Networks

RNNs address this limitation by maintaining an internal hidden state that implicitly encodes relevant history, removing the need to explicitly specify a fixed lag order. In canonical state-space form [Danilo P Mandic and Chambers, 2001], the state and output at time n are governed by

$$\mathbf{s}[n] = \boldsymbol{\varphi}(\mathbf{s}[n-1], \mathbf{u}[n-1], \hat{\mathbf{x}}[n-1]), \quad (2.21)$$

$$\hat{\mathbf{x}}[n] = \boldsymbol{\psi}(\mathbf{s}[n-1], \mathbf{u}[n-1], \hat{\mathbf{x}}[n-1]), \quad (2.22)$$

where $\mathbf{s}[n] \in \mathbb{R}^{N_s}$ is the hidden state vector of dimension N_s , $\mathbf{u}[n] \in \mathbb{R}^{N_u}$ is the external input vector of dimension N_u , $\hat{\mathbf{x}}[n] \in \mathbb{R}^N$ is the predicted output, and $\boldsymbol{\varphi}, \boldsymbol{\psi}$ are learnable non-linear functions for the state transition and output mapping, respectively. The recurrence in (2.21) allows the state to accumulate information across arbitrarily many time steps, in principle giving the network access to the entire input history without an explicit window length.

The simplest instantiation, the Elman network [Elman, 1990], simplifies (2.21)–(2.22) by removing output feedback and external input dependencies, yielding

$$\mathbf{h}[n] = \boldsymbol{\sigma}(\mathbf{W}_h \mathbf{h}[n-1] + \mathbf{W}_u \mathbf{u}[n] + \mathbf{b}_h), \quad (2.23)$$

$$\hat{\mathbf{x}}[n] = \mathbf{W}_o \mathbf{h}[n] + \mathbf{b}_o, \quad (2.24)$$

where $\boldsymbol{\sigma}(\cdot)$ is an element-wise non-linearity (typically tanh or ReLU), $\mathbf{W}_h, \mathbf{W}_u, \mathbf{W}_o$ are learnable weight matrices, and $\mathbf{b}_h, \mathbf{b}_o$ are bias vectors.

From a signal processing perspective, RNNs can be interpreted as adaptive non-linear IIR filters [Danilo P Mandic, 2002]: the feedback through $\mathbf{h}[n-1]$ creates an infinite impulse response, while the learned non-linearities enable modelling of complex temporal patterns beyond linear methods. However, standard RNNs suffer from vanishing and exploding gradients during training [Bengio, Simard, and Frasconi, 1994], limiting their ability to capture long-range dependencies. This motivated gated architectures that regulate information flow through the hidden state, most notably the long short-term memory (LSTM) [Hochreiter and Schmidhuber, 1997] and the gated recurrent unit (GRU) [Cho, Van Merriënboer, et al., 2014]. More recent architectures, including Transformers [Vaswani et al., 2017] and structured state-space models [A. Gu, Goel, and Re, 2022], aim to model sequential dependencies with improved scalability. These temporal sequence models serve as components within the spatiotemporal graph architectures discussed in Sec. 2.3.2, where they are coupled with graph neural networks to jointly model temporal evolution and spatial structure.

Remark 6. *The sequence models discussed above commonly produce deterministic point predictions. In many applications, however, quantifying predictive uncertainty is as important as the prediction itself. Probabilistic extensions of neural sequence models address this by parameterising predictive distributions rather than single outputs: examples include quantile regression networks that estimate conditional quantiles [R. Wen et al., 2017; Benidis et al., 2022], and score-based generative models such as denoising diffusion probabilistic models (DDPMs) that learn to sample from the full conditional distribution over missing or future values [J. Ho, Jain, and Abbeel, 2020; Tashiro et al., 2021; Alcaraz and Strodthoff, 2023; Jenkins, Z. Chen, et al., 2023]. These probabilistic formulations are revisited in Chapters 4 and 6, where they underpin template-guided imputation and calibrated uncertainty estimation, respectively.*

2.2 Signals and Systems on Irregular Domains: Graph Signal Processing

Traditional signal processing theory has been developed predominantly for signals defined on regular, structured domains with natural notions of shift-invariance and frequency. However, many modern datasets exhibit irregular structure: sensor networks, biological networks, transportation systems, and molecular structures all naturally reside on *graphs* rather than regular *grids*. Formulating signal processing on graphs enables principled treatment of signals on irregular domains, incorporates relational structure as an inductive bias, and generalises classical operations such as filtering and spectral analysis to these domains [Stanković, D. Mandić, et al., 2020b].

This section outlines the foundational theory of signal processing on graphs, extending the discrete-time concepts from Sec. 2.1 to irregular network structures. The discussion begins with graphs and signals (Sec. 2.2.1), before introducing the graph shift operator (Sec. 2.2.2) and graph filtering (Sec. 2.2.3), which generalise the temporal shift operator and convolution to irregular domains. Spectral graph theory is developed in Sec. 2.2.4, establishing the graph Laplacian, graph Fourier transform, and spectral filtering that provide the frequency-domain perspective for graph signals. Finally, graph machine learning (Sec. 2.2.5) integrates these filtering operations with learnable non-linear transformations, from spectral graph convolutional networks to spatial message passing architectures. These graph signal processing concepts, combined with the temporal methods from Sec. 2.1, provide the foundation for graph-based spatiotemporal machine learning (Sec. 2.3).

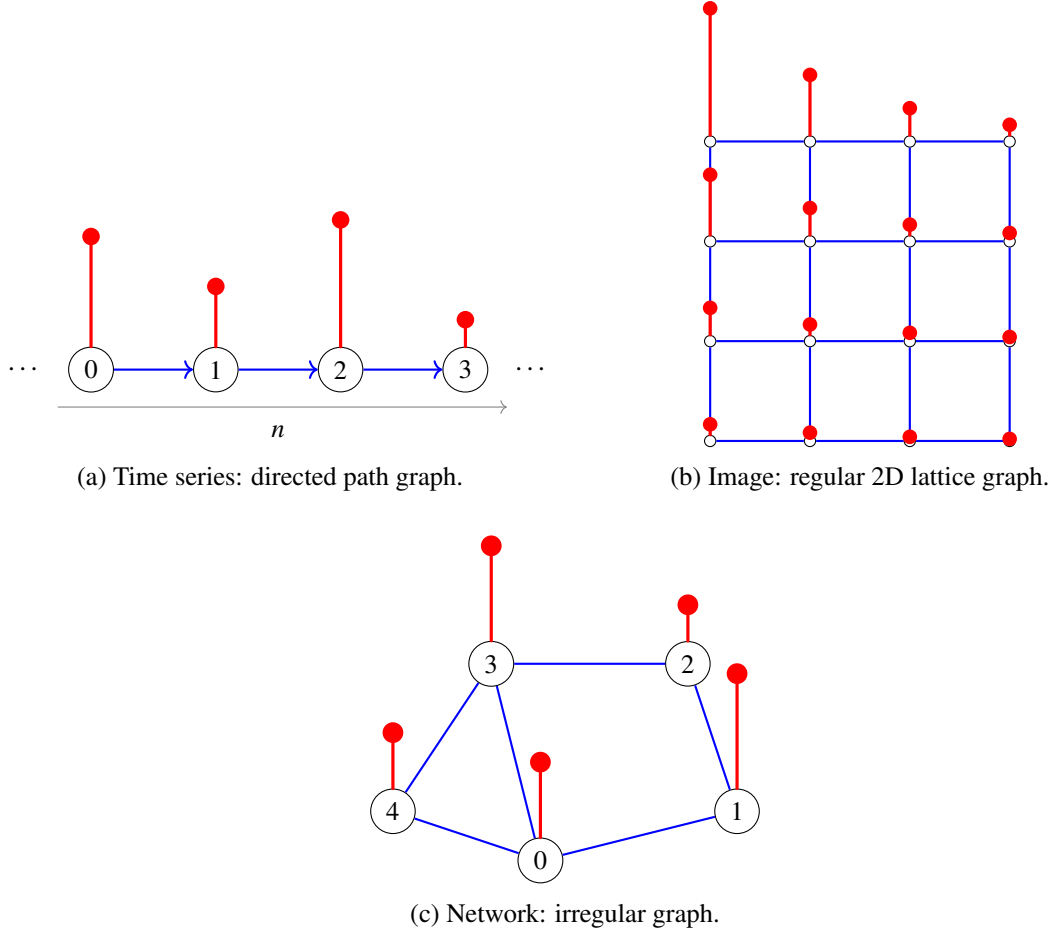


Figure 2.4: Signal domains as graphs of increasing structural irregularity. (a) A discrete-time signal resides on a directed path graph, where each node corresponds to a time index and edges encode the causal temporal ordering. (b) An image is naturally represented on a regular 2D lattice, where each pixel is a node connected to its spatial neighbours. (c) Many real-world systems exhibit irregular connectivity that cannot be captured by regular structures. Red stems indicate graph signal values at each node. Classical signal processing addresses domains (a) and (b); graph signal processing generalises these operations to arbitrary topologies such as (c).

2.2.1 Graphs and Signals

Definition of a Graph

A graph, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, is a mathematical structure representing relationships between entities (see Fig. 2.4c), where $\mathcal{V} = \{0, \dots, N-1\}$ denotes a set of N nodes and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ represents the set of edges connecting them. The connectivity structure is encoded in an adjacency matrix, $\mathbf{A} \in \mathbb{R}^{N \times N}$, where each entry, A_{ij} , indicates the presence of an edge from node i to node j . For weighted graphs, a weight matrix $\mathbf{W} \in \mathbb{R}^{N \times N}$ is employed, where W_{ij} quantifies the strength of the connection between nodes i and j . Elements W_{ij} are non-zero if and only if $(i, j) \in \mathcal{E}$.

For undirected graphs, the adjacency and weight matrices satisfy $A_{ij} = A_{ji}$ and $W_{ij} = W_{ji}$, respectively. Conversely, directed graphs permit asymmetric relationships where $W_{ij} \neq W_{ji}$ in general.

The degree of node i is defined as $d_i = \sum_{j=0}^{N-1} A_{ij}$ for unweighted graphs, or $d_i = \sum_{j=0}^{N-1} W_{ij}$ for weighted graphs. These degrees are collected in the diagonal degree matrix, $\mathbf{D} \in \mathbb{R}^{N \times N}$, where $D_{ii} = d_i$ and $D_{ij} = 0$ for $i \neq j$.

Signals on Graphs

A graph signal associates a value with each node in the graph, extending the concept of discrete-time signals from regularly indexed sequences to irregular network structures (see Fig. 2.4). Formally, a scalar graph signal is a function $x: \mathcal{V} \rightarrow \mathbb{R}$ that assigns a real value to each vertex. While the discussion focuses on scalar, real-valued signals for clarity, the framework naturally extends to complex-valued or vector-valued signals when multiple features are observed at each node.

The finite vertex set, $\mathcal{V}$, allows any graph signal to be represented equivalently as an N -dimensional vector, $\mathbf{x} = [x_0, \dots, x_{N-1}]^\top \in \mathbb{R}^N$, where x_i denotes the signal value at vertex v_i . This vector representation provides the foundation for linear algebraic operations on graph signals, analogous to how discrete-time signals are represented as vectors indexed by time. Throughout this section, $\mathbf{x} \in \mathbb{R}^N$ denotes a static graph signal indexed over nodes rather than the temporal regressor of Sec. 2.1.4; when signals evolve in time, the notation $\mathbf{x}[t]$ refers to the graph signal at time step t (Sec. 2.3).

2.2.2 The Graph Shift Operator

The graph shift operator extends the fundamental concept of temporal delay to irregular network structures, as illustrated in Fig. 2.5. In classical DSP, the time shift operator, z^{-1} (Eq. 2.1.1), delays a signal by one time step, replacing each sample, $x[n]$, with $x[n-1]$. On a graph, there is no natural notion of ‘previous’ or ‘next’ in time; instead, the shift operation propagates signal values along edges, from each node to its neighbours.

The graph shift operator, $\mathbf{S}$, is an $N \times N$ matrix that implements this propagation, defined such that the shifted signal, $\tilde{\mathbf{x}} = \mathbf{S}\mathbf{x}$, satisfies

$$\tilde{x}_i = [\mathbf{S}\mathbf{x}]_i = \sum_{j \in \mathcal{N}_i} S_{ij} x_j, \quad (2.25)$$

where $\mathcal{N}_i$ denotes the neighbourhood of vertex v_i and S_{ij} is the (i, j) -th entry of $\mathbf{S}$. This operation replaces the signal value at each node with a weighted combination of values from its graph neighbours.

A natural choice for the graph shift operator is the adjacency matrix, $\mathbf{S} = \mathbf{A}$, where $A_{ij} = W_{ij}$ if

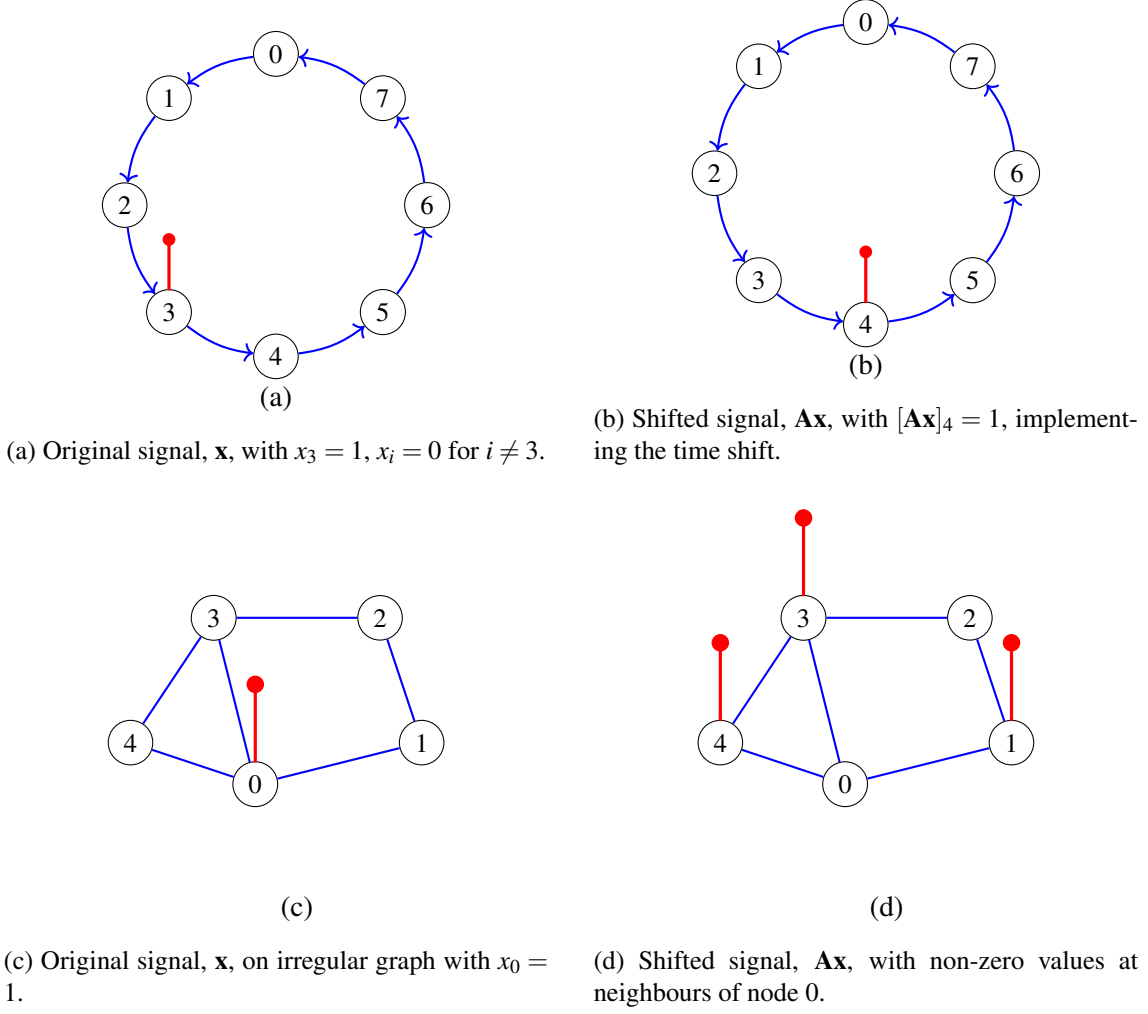


Figure 2.5: Graph shift operator on regular and irregular topologies. (a-b) On a directed circular graph representing periodic time series, the graph shift, $\mathbf{Ax}$, implements the classical time shift, moving the impulse from node 3 to node 4. (c-d) On an irregular graph, the shift diffuses the impulse at node 0 to all its neighbours (nodes 1, 3, and 4), demonstrating spatial propagation along graph edges.

$(j, i) \in \mathcal{E}$ and zero otherwise. The shift operation (2.25) then becomes

$$\tilde{\mathbf{x}} = \mathbf{Ax}. \quad (2.26)$$

Just as repeated application of the temporal shift yields multi-step delays, $(z^{-k}x)[n] = x[n - k]$, repeated application of the graph shift propagates information along paths of increasing length: $[\mathbf{S}^k \mathbf{x}]_i$ aggregates signal values from all nodes reachable from v_i via walks of length exactly k .

Remark 7. Alternative choices for $\mathbf{S}$ include the Laplacian matrix, $\mathbf{L} = \mathbf{D} - \mathbf{A}$, or normalised variants. The choice of shift operator determines the notion of signal propagation on the graph, the spectral properties, and the localisation behaviour of graph filters. Unlike the temporal shift, which is unitary and bijective, the graph shift is generally a many-to-one operation that does not

preserve signal energy. This has motivated the development of isometric [Girault, Gonçalves, and Fleury, 2015], unitary [Dees et al., 2019], and doubly stochastic shift operators [Scalzo et al., 2023] that restore desirable properties of the classical time shift to the graph domain.

2.2.3 Graph Filters

Analogous to discrete-time filtering via convolution (Sec. 2.1.1), graph signal processing employs filters to process signals on graphs. A graph filter, $\mathcal{F}$, is a system that maps an input graph signal, $\mathbf{x} \in \mathbb{R}^N$, to an output graph signal, $\mathbf{y} \in \mathbb{R}^N$.

Linearity. A graph filter, $\mathcal{F}$, is linear if it satisfies the superposition principle: for any graph signals $\mathbf{x}', \mathbf{x}'' \in \mathbb{R}^N$ and scalars a_1, a_2 ,

$$\mathcal{F}(a_1\mathbf{x}' + a_2\mathbf{x}'') = a_1\mathcal{F}(\mathbf{x}') + a_2\mathcal{F}(\mathbf{x}''). \quad (2.27)$$

Linear graph filters can be represented as matrix multiplication, $\mathbf{y} = \mathbf{F}\mathbf{x}$, where $\mathbf{F} \in \mathbb{R}^{N \times N}$ is the filter matrix realisation of the operator $\mathcal{F}$.

Shift-Invariance. A linear graph filter is shift-invariant if it commutes with the graph shift operator,

$$\mathbf{F}\mathbf{S} = \mathbf{S}\mathbf{F}, \quad (2.28)$$

that is, filtering a shifted signal is equivalent to shifting the filtered signal.

Polynomial Graph Filters. Any linear, shift-invariant graph filter can be expressed as a polynomial in $\mathbf{S}$, making polynomial graph filters the complete class of linear, shift-invariant filters on graphs [Sandryhaila and Moura, 2013]. A polynomial graph filter of order K has matrix realisation

$$\mathbf{F}_K = \sum_{k=0}^K h_k \mathbf{S}^k = h_0 \mathbf{I} + h_1 \mathbf{S} + \dots + h_K \mathbf{S}^K, \quad (2.29)$$

where $\mathbf{h} = [h_0, \dots, h_K]^\top$ is the vector of filter coefficients. Fig. 2.6 illustrates this decomposition for $K=2$, showing how each term $h_k \mathbf{S}^k \mathbf{x}$ aggregates signal values from the k -hop neighbourhood. The filter order, K , determines the spatial support: information propagates at most K hops from each node, analogous to the temporal support of FIR filters.

Remark 8. The polynomial representation (2.29) parallels temporal filtering, where filters are polynomials in z^{-1} (Eq. 2.1.1), with $\mathbf{S}^k$ propagating information across k -hop neighbourhoods analogously to k -step delays. While the filter coefficients $\mathbf{h}$ are applied uniformly (making the filter

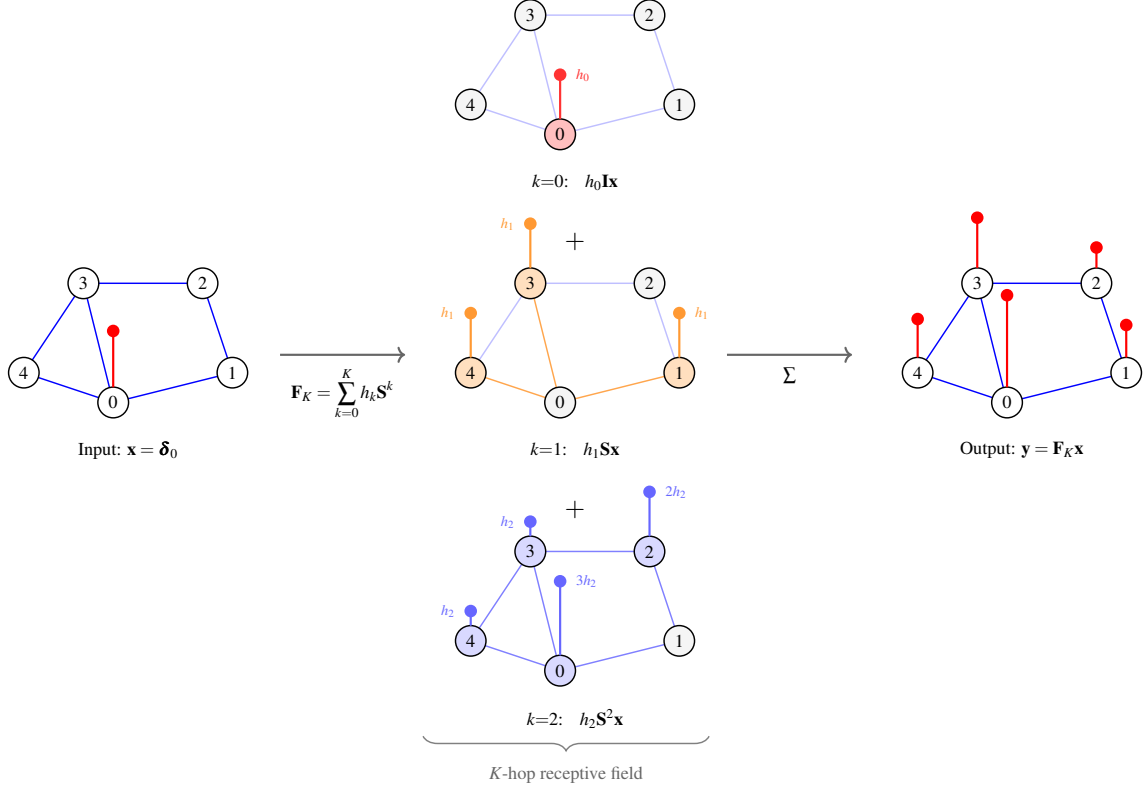


Figure 2.6: Polynomial graph filter of order $K=2$ applied to an impulse, $\mathbf{x} = \delta_0$, on the 5-node graph from Fig. 2.5. Each term, $h_k \mathbf{S}^k \mathbf{x}$, contributes signal values at nodes reachable via walks of exactly k hops from the source: $k=0$ (self, red), $k=1$ (immediate neighbours, orange), $k=2$ (2-hop walks, blue). The stem heights reflect the number of length- k walks from node 0 to each node, for instance, three length-2 walks return to node 0 while none reach node 1. This illustrates how graph topology determines the filter’s effective impulse response. Unweighted adjacency ($\mathbf{S} = \mathbf{A}$) is assumed.

shift-invariant), the heterogeneous structure of graphs means this does not correspond to classical convolution: nodes with different neighbourhood sizes aggregate different numbers of values even with identical coefficients. The spectral perspective (Sec. 2.2.4) provides an alternative view that more directly parallels the frequency-domain interpretation of temporal convolution.

2.2.4 Spectral Graph Theory

Classical signal processing leverages the Fourier transform to analyse signals in the frequency domain, where convolution becomes pointwise multiplication. Spectral graph theory extends this frequency-domain perspective to graphs through the eigendecomposition of the graph shift operator, providing a rigorous notion of frequency for signals defined on irregular domains. Just as polynomial graph filters (Sec. 2.2.3) generalise temporal FIR filters by replacing the time-shift operator, z^{-1} , with the graph shift operator, $\mathbf{S}$, the spectral framework generalises classical Fourier analysis by replacing the complex exponentials of the discrete Fourier transform with the eigenvectors of $\mathbf{S}$.

The Graph Laplacian. For an undirected weighted graph, $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{W})$, the graph Laplacian is defined as

$$\mathbf{L} = \mathbf{D} - \mathbf{W}, \quad (2.30)$$

where $\mathbf{D} \in \mathbb{R}^{N \times N}$ is the diagonal degree matrix with $D_{ii} = \sum_{j=0}^{N-1} W_{ij}$. For unweighted graphs, $\mathbf{W}$ reduces to the adjacency matrix, $\mathbf{A}$, and $D_{ii} = d_i$. The Laplacian acts as a difference operator on graph signals, where the i -th component of $\mathbf{L}\mathbf{x}$ computes a weighted sum of differences between the signal value at node i and those at its neighbours,

$$[\mathbf{L}\mathbf{x}]_i = \sum_{(i,j) \in \mathcal{E}} W_{ij}(x_i - x_j). \quad (2.31)$$

This interpretation motivates the Laplacian as a natural choice of graph shift operator, $\mathbf{S}$, when signal smoothness with respect to the graph topology is of primary interest. Alternatively, the normalised graph Laplacian is given by

$$\mathbf{L}_{\text{norm}} = \mathbf{I}_N - \mathbf{D}^{-1/2} \mathbf{W} \mathbf{D}^{-1/2}, \quad (2.32)$$

which accounts for heterogeneous node degrees and has eigenvalues bounded in $[0, 2]$ [Chung, 1997].

As $\mathbf{L}$ is real symmetric and positive semi-definite, it admits an eigendecomposition

$$\mathbf{L} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^\top, \quad (2.33)$$

where $\mathbf{U} = [\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_{N-1}] \in \mathbb{R}^{N \times N}$ contains the orthonormal eigenvectors and $\mathbf{\Lambda} = \text{diag}(\lambda_0, \lambda_1, \dots, \lambda_{N-1})$ contains the corresponding non-negative eigenvalues, ordered $0 = \lambda_0 \leq \lambda_1 \leq \dots \leq \lambda_{N-1}$. The smallest eigenvalue $\lambda_0 = 0$ always exists, with corresponding eigenvector $\mathbf{u}_0 = \frac{1}{\sqrt{N}} \mathbf{1}$ (the constant signal), and the multiplicity of the zero eigenvalue equals the number of connected components in $\mathcal{G}$.

The Graph Fourier Transform. The eigenvectors, $\{\mathbf{u}_k\}_{k=0}^{N-1}$, form an orthonormal basis for $\mathbb{R}^N$ and serve as the Fourier basis for graph signals. The GFT of a graph signal, $\mathbf{x}$, is given by

$$\hat{\mathbf{x}} = \mathbf{U}^\top \mathbf{x}, \quad (2.34)$$

where $\hat{x}_k = \mathbf{u}_k^\top \mathbf{x}$ is the spectral coefficient associated with graph frequency λ_k . The inverse graph Fourier transform (IGFT) reconstructs the signal as a superposition of the Fourier modes,

$$\mathbf{x} = \mathbf{U}\hat{\mathbf{x}} = \sum_{k=0}^{N-1} \hat{x}_k \mathbf{u}_k. \quad (2.35)$$

This parallels the classical discrete Fourier transform, where signals are decomposed into complex exponentials of increasing frequency; here, graph signals are decomposed into eigenvectors of increasing graph frequency λ_k .

Graph Frequencies and Smoothness. The eigenvalues $\{\lambda_k\}$ carry a natural frequency interpretation grounded in signal smoothness [Chung, 1997]. The total variation of a graph signal, $\mathbf{x}$, with respect to the graph topology is quantified by the Laplacian quadratic form,

$$\mathbf{x}^\top \mathbf{L} \mathbf{x} = \frac{1}{2} \sum_{(i,j) \in \mathcal{E}} W_{ij} (x_i - x_j)^2, \quad (2.36)$$

which measures the aggregate squared difference between signal values at connected nodes. Since the eigenvectors are orthonormal, evaluating (2.36) on the k -th eigenvector yields $\mathbf{u}_k^\top \mathbf{L} \mathbf{u}_k = \lambda_k$. Eigenvectors with small eigenvalues therefore vary slowly across edges (adjacent nodes carry similar values), while eigenvectors with large eigenvalues exhibit rapid oscillation (large differences between connected nodes). This provides a natural frequency ordering on the graph: small λ_k correspond to low-frequency (smooth) components and large λ_k to high-frequency (rapidly varying) components, directly analogous to the ordering of classical Fourier frequencies.

Spectral Filtering. Graph filtering can be defined in the spectral domain by modulating each frequency component independently. A spectral filter with frequency response, $g(\cdot)$, applied to a graph signal, $\mathbf{x}$, yields

$$\mathbf{y} = g(\mathbf{L})\mathbf{x} = \mathbf{U}g(\mathbf{\Lambda})\mathbf{U}^\top \mathbf{x}, \quad (2.37)$$

where $g(\mathbf{\Lambda}) = \text{diag}(g(\lambda_0), \dots, g(\lambda_{N-1}))$ applies the filter's frequency response to each spectral component. This operation corresponds to three steps: (i) transform to the spectral domain via the GFT, (ii) multiply each frequency component by the corresponding filter gain, $g(\lambda_k)$, and (iii) transform back via the IGFT. For example, $g(\lambda) = \mathbb{1}[\lambda \leq \lambda_c]$ defines a low-pass graph filter that retains spectral components below the cut-off frequency, λ_c , while $g(\lambda) = \mathbb{1}[\lambda > \lambda_c]$ defines the corresponding high-pass filter. Here, $\mathbb{1}[\cdot]$ denotes the indicator function, equal to one when its argument holds and zero otherwise.

For polynomial frequency responses, $g(\lambda) = \sum_{p=0}^K h_p \lambda^p$, the spectral filtering operation (2.37)

admits an efficient vertex-domain implementation. Substituting the eigendecomposition, $\mathbf{L} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$,

$$\mathbf{y} = \mathbf{U}g(\mathbf{\Lambda})\mathbf{U}^\top\mathbf{x} = \mathbf{U}\left(\sum_{p=0}^K h_p\mathbf{\Lambda}^p\right)\mathbf{U}^\top\mathbf{x} = \sum_{p=0}^K h_p\mathbf{U}\mathbf{\Lambda}^p\mathbf{U}^\top\mathbf{x} = \sum_{p=0}^K h_p\mathbf{L}^p\mathbf{x}, \quad (2.38)$$

where the final equality follows from $\mathbf{U}\mathbf{\Lambda}^p\mathbf{U}^\top = (\mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top)^p = \mathbf{L}^p$. This vertex-domain representation (2.38) can be computed efficiently using K sparse matrix-vector multiplications at cost $\mathcal{O}(K|\mathcal{E}|)$, entirely avoiding the $\mathcal{O}(N^2)$ complexity of multiplying by the Fourier basis, $\mathbf{U}$. Despite being computed in the vertex domain, polynomial graph filters achieve the same spectral filtering effect as (2.37); they implicitly apply the frequency response, $g(\lambda_k)$, to each spectral component.

Remark 9. *The Laplacian-based spectral framework provides a well-established foundation for graph signal processing with a clear frequency interpretation: small eigenvalues correspond to smooth, slowly varying signals, and large eigenvalues to rapidly oscillating ones. However, the Laplacian is restricted to undirected graphs with non-negative edge weights, and its eigendecomposition changes with the graph topology, meaning that filters learned on one graph do not directly transfer to another. Convolution through the graph Fourier transform has been generalised to directed and signed graphs using the adjacency matrix as the shift operator [Sandryhaila and Moura, 2013; Shafipour et al., 2018], though the resulting spectral domains may lack the smoothness-based frequency interpretation of the Laplacian. These spectral concepts underpin the graph convolutional networks developed in Sec. 2.2.5.*

2.2.5 Graph Machine Learning

The graph filters and spectral methods of the preceding sections provide principled tools for processing signals on graphs, but they are inherently linear and operate with fixed, (possibly hand-designed) coefficients. GNNs extend this framework by introducing learnable, non-linear transformations of graph-structured data while respecting the relational inductive biases encoded in the graph topology [Scarselli et al., 2008; Bacciu et al., 2020; Bronstein, Bruna, Cohen, et al., 2021].

Early forms of neural processing on graph-structured data can be traced to the 1990s, with work on recursive architectures for structured data (such as directed acyclic graphs or trees) [Sperduti, 1993; Goller and Kuchler, 1996; Sperduti and Starita, 1997; Frasconi, Gori, and Sperduti, 1998]. The term graph neural network was introduced by Gori, Monfardini, and Scarselli (2005) and Scarselli et al. (2008), who proposed a recurrent framework computing node representations via fixed-point iteration of contraction mappings, while Micheli (2009) concurrently developed a feed-forward alternative. The Gated Graph Neural Network of Yujia Li et al. (2015) brought modern recurrent mechanisms to the GNN model, including gating and back-propagation through

time. However, the architectures most widely used today emerged from applying spectral graph theory to define learnable convolution on graphs [Bruna et al., 2014; Defferrard, Bresson, and Vandergheynst, 2016; Kipf and Welling, 2017], and their subsequent reinterpretation through the spatial lens of message passing [Gilmer et al., 2017; Battaglia et al., 2018]. This section reviews both perspectives, which directly extend the graph filtering framework of Sec. 2.2.3 to the learnable, non-linear setting.

Spectral Graph Convolutional Networks

The spectral approach to GNNs builds on the GFT framework of Sec. 2.2.4, making the filter’s frequency response a learnable function optimised via back-propagation. The first spectral graph convolution network (GCN) [Bruna et al., 2014] parameterises the frequency response directly, replacing the fixed $g(\mathbf{\Lambda})$ in (2.37) with a learnable diagonal. For a single-channel graph signal $\mathbf{x}^{(l)} \in \mathbb{R}^N$ at layer l , the filtered output is

$$\mathbf{x}^{(l+1)} = \sigma \left(\mathbf{U} \mathbf{\Theta}^{(l)} \mathbf{U}^\top \mathbf{x}^{(l)} \right), \quad (2.39)$$

where $\mathbf{\Theta}^{(l)} = \text{diag}(\theta_0^{(l)}, \dots, \theta_{N-1}^{(l)})$ contains learnable spectral coefficients and $\sigma(\cdot)$ is a pointwise non-linearity. The product, $\mathbf{U} \mathbf{\Theta}^{(l)} \mathbf{U}^\top$, applies a learnable frequency response in the graph spectral domain, directly generalising the fixed spectral filters $g(\mathbf{\Lambda})$ of (2.37) to the trainable setting. The extension to multi-channel node features, where each node carries c_l input and c_{l+1} output channels, replaces the single filter with a bank of $c_l \times c_{l+1}$ learnable spectral filters, one per input–output channel pair. However, this formulation requires $\mathcal{O}(N)$ parameters per channel pair, the forward and inverse GFT each cost $\mathcal{O}(N^2)$, and the learned coefficients are tied to the eigenbasis, $\mathbf{U}$, of a specific graph, preventing transfer across topologies. Moreover, because localisation in the vertex domain arises only indirectly through smoothness of the learned frequency response, the formulation does not provide explicit control over the spatial support of filters [Bruna et al., 2014].

These limitations are addressed by constraining the frequency response to a polynomial. Recalling from (2.38) that a polynomial frequency response, $g(\lambda) = \sum_{p=0}^K h_p \lambda^p$, can be realised in the vertex-domain as $\sum_{p=0}^K h_p \mathbf{L}^p \mathbf{x}$, Defferrard, Bresson, and Vandergheynst (2016) proposed *ChebNet*, which parameterises the frequency response using Chebyshev polynomials of the first kind,

$$g_{\boldsymbol{\theta}}(\mathbf{L})\mathbf{x} = \sum_{k=0}^K \theta_k T_k(\tilde{\mathbf{L}}) \mathbf{x}, \quad (2.40)$$

where $T_k(\cdot)$ is the k -th Chebyshev polynomial, $\tilde{\mathbf{L}} = 2\mathbf{L}/\lambda_{\max} - \mathbf{I}_N$ rescales the spectrum to $[-1, 1]$, and $\boldsymbol{\theta} = [\theta_0, \dots, \theta_K]^\top$ are the learnable coefficients. The Chebyshev recurrence, $T_k(x) = 2xT_{k-1}(x) -$

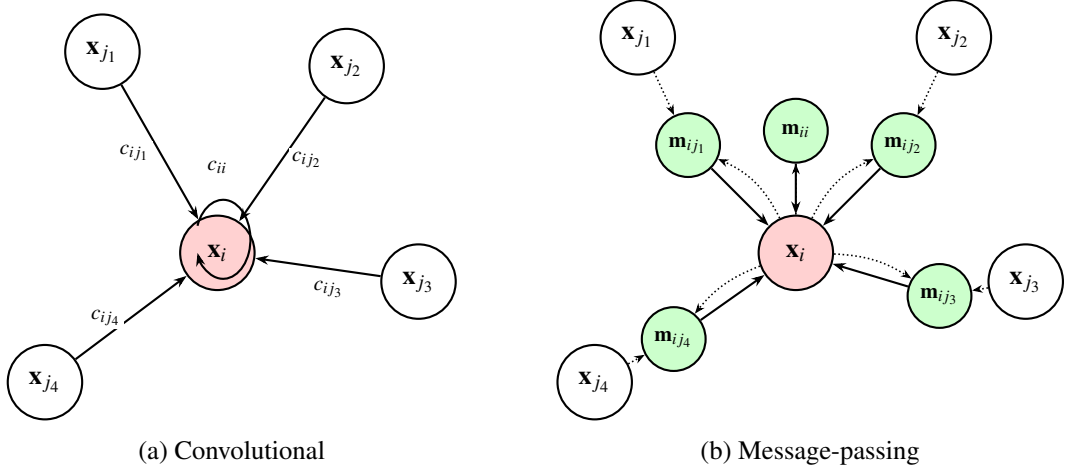


Figure 2.7: Single-layer update at node i under convolutional and message-passing aggregation from neighbours ($j_1, \dots, j_4 \in \mathcal{N}_i$). (a) Neighbouring node features, $\mathbf{x}_{j_k}$, are scaled by fixed, topology-dependent coefficients c_{ij_k} (e.g. normalised edge weights) and aggregated directly at node i (solid arrows), along with a self-contribution, $c_{ii}\mathbf{x}_i$. (b) Each edge produces a learned message, $\mathbf{m}_{ij_k} = \rho(\mathbf{x}_i, \mathbf{x}_{j_k}, \mathbf{e}_{ij_k})$, that depends on both endpoint features (dotted arrows into green nodes), before aggregation at node i (solid arrows). Since the message function has access to both $\mathbf{x}_i$ and $\mathbf{x}_{j_k}$, the aggregation is anisotropic, meaning different neighbours contribute qualitatively different information, unlike the isotropic weighting of convolutional methods.

$T_{k-2}(x)$, allows each term to be computed iteratively via sparse matrix-vector multiplications at cost $O(K|\mathcal{E}|)$, entirely avoiding the eigendecomposition. The polynomial order K directly controls the spatial support, providing K -hop localisation that the unconstrained spectral approach of (2.39) lacks.

Kipf and Welling (2017) further simplified ChebNet by restricting to $K = 1$, yielding the widely-adopted GCN layer,

$$\mathbf{X}^{(l+1)} = \sigma\left(\tilde{\mathbf{D}}^{-1/2} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-1/2} \mathbf{X}^{(l)} \mathbf{W}^{(l)}\right), \quad (2.41)$$

where $\mathbf{W}^{(l)} \in \mathbb{R}^{d_l \times d_{l+1}}$ are learnable weights, $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}_N$ is the augmented adjacency with self-connections, and $\tilde{\mathbf{D}}$ is the corresponding degree matrix. Each layer amounts to a normalised aggregation of features in the 1-hop neighbourhood, a linear transformation, and a non-linearity. The normalised adjacency multiplication computes a weighted average of each node's features with those of its neighbours, which essentially acts as a low-pass graph filter in the sense of Sec. 2.2.4. At this level of simplification, the spectral and spatial interpretations converge: the GCN layer is equivalently a first-order polynomial graph filter ((2.29) with $K=1$ and $\mathbf{S} = \tilde{\mathbf{D}}^{-1/2} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-1/2}$) composed with a learnable channel mixing and pointwise non-linearity. Deep architectures are constructed by stacking L such layers, yielding an L -hop receptive field in which interleaved non-linearities enable the network to learn representations that no linear polynomial graph filter can express.

Spatial Graph Neural Networks

While the spectral perspective provides rigorous mathematical foundations, it is restricted to undirected graphs with a symmetric Laplacian and ties filter design to a specific eigen-basis. The spatial perspective instead defines GNN layers directly in terms of operations on the graph neighbourhood, without reference to spectral decomposition. The message-passing neural network (MPNN) framework [Gilmer et al., 2017; Battaglia et al., 2018] formalises this view and is illustrated in Fig. 2.7.

At layer l , the representation of node i is updated through three operations. First, computing a message along each incoming edge; second, aggregating these messages from all neighbours; and third, updating the node state. In its most general form, the message passing formulation is written as

$$\mathbf{x}_i^{(l+1)} = \gamma^{(l)} \left(\mathbf{x}_i^{(l)}, \bigoplus_{j \in \mathcal{N}_i} \rho^{(l)}(\mathbf{x}_i^{(l)}, \mathbf{x}_j^{(l)}, \mathbf{e}_{ij}) \right), \quad (2.42)$$

where $\rho^{(l)}(\cdot)$ is a learnable message function computing a vector sent from node j to node i , $\bigoplus$ is a permutation-invariant aggregation function (such as summation, mean, or max), $\gamma^{(l)}(\cdot)$ is a learnable update function, and $\mathbf{e}_{ij}$ denotes optional edge attributes. Both $\rho^{(l)}$ and $\gamma^{(l)}$ can be implemented as arbitrary differentiable functions, such as an multilayer perceptron (MLP) or gated mechanism [Gilmer et al., 2017].

Many well-known architectures are recovered as special cases of (2.42) through increasingly restrictive choices of $\rho^{(l)}$. For instance, for convolutional layers [Kipf and Welling, 2017], the message reduces to $\rho^{(l)}(\mathbf{x}_i^{(l)}, \mathbf{x}_j^{(l)}, \mathbf{e}_{ij}) = c_{ij} \mathbf{W}^{(l)} \mathbf{x}_j^{(l)}$, where c_{ij} is a fixed scalar determined by the graph structure (e.g. normalised edge weight) and the same weight matrix $\mathbf{W}^{(l)}$ is applied to every neighbour. The aggregation is therefore isotropic, meaning the variation across neighbours arises only through the structural coefficients c_{ij} , directly paralleling the graph filters of Sec. 2.2.3. In attentional layers [Veličković et al., 2018], the fixed coefficients c_{ij} are replaced by learned, data-dependent attention weights $a_{ij}^{(l)}$ computed from both sender and receiver node features, making the aggregation anisotropic while keeping the message itself a linear function of the sender. In the general message-passing case, $\rho^{(l)}$ computes arbitrary non-linear functions of both node representations and edge attributes. These three formulations form a representational hierarchy [Bronstein, Bruna, Cohen, et al., 2021],

$$\text{convolutional} \subseteq \text{attentional} \subseteq \text{message-passing}, \quad (2.43)$$

since any convolutional layer can be written as an attentional layer (with fixed attention weights),

and any attentional layer can be written as a message-passing layer (with scalar-weighted linear messages). However, the reverse does not hold, general message-passing layers can compute functions that neither attentional nor convolutional layers can represent.

Remark 10. *The message-passing formulation in (2.42) provides the framework used throughout the results chapters. In Chapter 5, $\rho^{(l)}$ is implemented as an MLP on concatenated sender and receiver features with learned scalar edge weighting, and $\gamma^{(l)}$ incorporates residual skip connections. In Chapter 6, message-passing layers process prediction residuals over a sparse learned graph topology. GNNs share parameters across all neighbourhoods and stacking L layers expands the receptive field to L -hop neighbourhoods, in a manner analogous to increasing the order of polynomial graph filters. However, excessive depth can cause over-smoothing, where representations become indistinguishable [Q. Li, Z. Han, and X.-M. Wu, 2018], and over-squashing, where information from distant nodes is compressed through topological bottlenecks [Alon and Yahav, 2021]. Further architectural developments, such as graph pooling, continuous-depth formulations, and higher-order extensions, are omitted here for brevity but discussed in subsequent chapters where relevant.*

2.3 Time-Varying Signals and Systems on Irregular Domains: Graph-Based Spatiotemporal Models

Many real-world phenomena are naturally described by a collection of time series whose evolution is governed not only by their individual temporal dynamics but also by cross-sectional relationships among them. A network of weather stations records temperature sequences that are spatially correlated through atmospheric transport; nodes in a social network generate activity streams shaped by community structure; and electrodes on the cardiac surface measure electrical potentials whose coupling reflects the tissue’s connectivity and conduction pathways.

In each case, the joint modelling of N coupled time series, $\{\mathbf{x}[t]\}_{t \in \mathbb{Z}}$, $\mathbf{x}[t] \in \mathbb{R}^N$, must contend with a fundamental dimensionality challenge: an unrestricted multivariate model, such as the VAR of Sec. 2.1.3, requires $\mathcal{O}(PN^2)$ parameters, rendering estimation intractable when N is large and when the number of temporal observations is smaller than the number of time series. Classical remedies include factor models [Stock and Watson, 2002; Forni et al., 2000], shrinkage estimators [Ledoit and M. Wolf, 2004; Yilun Chen, Wiesel, and Alfred O. Hero, 2011], and low-rank data representations [H.-F. Yu, N. Rao, and Dhillon, 2016].

This section focuses on a complementary and increasingly prominent approach: exploiting a graph that encodes the cross-sectional relationships among the time series to regularise and structure the

learning problem. The graph may reflect physical proximity (e.g. a sensor network embedded in Euclidean space), anatomical connectivity (e.g. cardiac tissue), or abstract relational structure (e.g. a social or financial network). Graph-based spatiotemporal models leverage this relational topology to constrain the cross-correlational structure of the signals, replacing dense parameter matrices with graph-structured operators whose complexity scales with graph sparsity rather than the square of the number of nodes. Graphs provide a flexible and unifying framework: they subsume regular domains (grids, lattices) as special cases while naturally accommodating irregular and heterogeneous topologies. This section begins with classical approaches that extend autoregressive models to graph-structured domains (Sec. 2.3.1), establishing the foundational connection between temporal and graph filters, before presenting a taxonomy of STGNNs organised by their architectures for spatial–temporal processing (Sec. 2.3.2).

2.3.1 Classical Spatiotemporal Models on Graphs

The foundational challenge in spatiotemporal modelling on graphs lies in extending classical time series analysis to graph-structured data. Consider a collection of N discrete-time signals evolving on a graph, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, with shift operator, $\mathbf{S}$. At each time step, t , the observed graph signal, $\mathbf{x}[t] \in \mathbb{R}^N$, yields a multivariate time series, $\{\mathbf{x}[t]\}_{t \in \mathbb{Z}}$. A classical VAR model (Sec. 2.1.3) would represent this as

$$\mathbf{x}[t] = \sum_{p=1}^P \mathbf{\Psi}_p \mathbf{x}[t-p] + \mathbf{e}[t], \quad (2.44)$$

where $\mathbf{\Psi}_p \in \mathbb{R}^{N \times N}$ are dense coefficient matrices and $\mathbf{e}[t]$ is white noise. This formulation requires learning $\mathcal{O}(PN^2)$ parameters, rendering it intractable for large graphs even with modest temporal orders, P . Moreover, the dense matrices, $\mathbf{\Psi}_p$, ignore the graph topology, $\mathcal{G}$, discarding structural information that could regularise parameter estimation [Mei and Moura, 2016].

Time-Vertex Stationarity and Filters

The key insight enabling tractable spatiotemporal models on graphs is the assumption of time-vertex stationarity, which generalises the notion of WSS from temporal signals (Sec. 2.1.3) to joint time–graph domains [Grassi et al., 2018; Isufi et al., 2019; Perraudin and Vandergheynst, 2017], where joint time–vertex Fourier transforms and filters are defined. A process of graph signals, $\{\mathbf{x}[t]\}$, is time-vertex stationary if its first- and second-order statistics are (i) invariant over

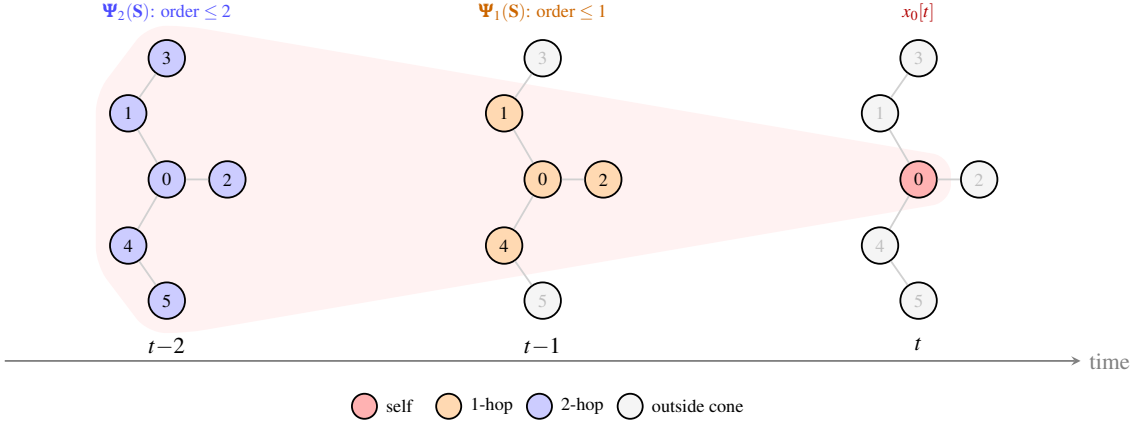


Figure 2.8: The ‘light cone’ of the causal graph process (CGP) model. To model the value of node 0 at time t (self, red), the CGP model constrains the graph filter order at each lag p to at most p : $\Psi_1(\mathbf{S})$ aggregates the 1-hop neighbourhood (orange), while $\Psi_2(\mathbf{S})$ reaches the 2-hop neighbourhood (blue). Nodes outside the cone (grey) cannot causally influence the prediction, enforcing a finite propagation speed of one graph shift per time step.

time and (ii) aligned with the graph topology. Formally, this requires

$$\mathbb{E}[\mathbf{x}[t]] = \boldsymbol{\mu}, \quad \forall t \in \mathbb{Z}, \quad (2.45)$$

$$\mathbb{E}[\mathbf{x}[t]\mathbf{x}[t-\tau]^\top] = \mathbf{R}[\tau] = \mathbf{U}\mathbf{r}_\tau(\boldsymbol{\Lambda})\mathbf{U}^\top, \quad \forall t, \tau \in \mathbb{Z}, \quad (2.46)$$

where $\mathbf{R}[\tau]$ is the lag- τ auto-covariance matrix, $\mathbf{U}$ and $\boldsymbol{\Lambda}$ are the eigenvectors and eigenvalues of the graph shift operator (Sec. 2.2.2), and $\mathbf{r}_\tau(\boldsymbol{\Lambda})$ is a diagonal matrix encoding temporal correlations in the graph spectral domain.

This stationarity assumption implies that spatial correlations align with the graph structure, while temporal correlations remain shift-invariant. Under this framework, the dense coefficient matrices, Ψ_p in (2.44), can be replaced by graph filters (Sec. 2.2.3), i.e. polynomials in the graph shift operator $\mathbf{S}$,

$$\Psi_p = \sum_{k=0}^{K_p} h_{p,k} \mathbf{S}^k, \quad (2.47)$$

where K_p is the filter order and $h_{p,k}$ are scalar coefficients. This representation reduces the number of parameters from $\mathcal{O}(N^2)$ to $\mathcal{O}(K_p)$ per lag, while ensuring that learned patterns respect the graph topology through localised operations (recall Sec. 2.2.3). The power $\mathbf{S}^k$ propagates information exactly k hops through the graph, providing explicit control over the spatial receptive field.

The Causal Graph Process

The graph VAR in (2.47) leaves the filter order K_p as a free parameter at each lag. A principled constraint, introduced by Mei and Moura (2016) as the causal graph process (CGP), couples the

spatial reach to the temporal lag by restricting $K_p \leq p$, meaning that the graph filter at lag p is a polynomial in $\mathbf{S}$ of order at most p . The resulting model is

$$\mathbf{x}[t] = \sum_{p=1}^P \Psi_p(\mathbf{S}) \mathbf{x}[t-p] + \mathbf{e}[t], \quad \Psi_p(\mathbf{S}) = \sum_{k=0}^p h_{p,k} \mathbf{S}^k, \quad (2.48)$$

where $h_{p,k}$ are scalar coefficients. The constraint $K_p \leq \min(p, N_{\mathbf{S}})$, where $N_{\mathbf{S}}$ is the degree of the minimal polynomial of $\mathbf{S}$, encodes a finite “propagation speed” on the graph: a signal emitted p time steps ago can influence at most nodes within a p -hop neighbourhood (see Fig. 2.8). Mei and Moura (2016) draw an analogy to the light cone in special relativity; nodes beyond the graph-distance horizon set by the temporal lag cannot causally influence the current observation. More general graph VAR formulations that include MA terms have also been developed [Isufi et al., 2019].

Remark 11. *Graph VAR models illustrate a fundamental trade-off in spatiotemporal modelling. The time-vertex stationarity assumption and graph filter parameterisation provide strong inductive biases that enable parameter-efficient learning from limited data, yielding interpretable models with explicit spatial receptive fields. However, these assumptions may be restrictive for systems exhibiting non-stationary non-linear dynamics, heterogeneous node behaviours, or time-varying graph topologies. Modern deep learning approaches, discussed in Sec. 2.3.2, relax these constraints through learned representations but sacrifice interpretability and require substantially more training data.*

2.3.2 Architectures for Spatiotemporal Graph Neural Networks

The classical graph models of Sec. 2.3.1 impose linearity and time-vertex stationarity. Modern STGNNs relax both constraints by coupling graph neural network layers (Fig. 2.2.5) with temporal sequence models (Eq. 2.1.5). The core design principle mirrors that of grid-based spatiotemporal architectures such as *ConvLSTM* [X. Shi et al., 2015]: replace spatial convolutions with graph message passing to extend local aggregation from regular grids to irregular topologies, while retaining temporal components, such as recurrent cells, temporal convolutions, or attention, to model the temporal dynamics. The graph acts as an inductive bias that regularises learning, particularly useful in data-scarce regimes, and enables transferability across graphs with different sizes or connectivity patterns [Bronstein, Bruna, LeCun, et al., 2017; Levie et al., 2021]. The central architectural decision concerns how spatial and temporal operations are coupled within each layer. The following taxonomy is organised by the order in which they are composed [Cini, Marisca, Zambon, et al., 2025; G. Jin et al., 2024].

Spatiotemporal Message-Passing

The basic computational block of an STGNN extends the spatial message-passing framework of (2.42) to aggregate information from each node’s spatiotemporal neighbourhood. In general, the underlying graph may evolve overtime (e.g. through changing connectivity or edge weights); for simplicity, a fixed graph topology, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, is considered throughout this section, noting that the framework extends naturally to the time-varying case.

A spatiotemporal message-passing (STMP) block at layer l that updates the representation of node i at time step t , as

$$\mathbf{h}_i^{(l+1)}[t] = \gamma_{\text{st}}^{(l)} \left(\mathbf{h}_i^{(l)}[\leq t], \bigoplus_{j \in \mathcal{N}_i} \rho_{\text{st}}^{(l)}(\mathbf{h}_i^{(l)}[\leq t], \mathbf{h}_j^{(l)}[\leq t], \mathbf{e}_{ji}) \right), \quad (2.49)$$

where $\mathcal{N}_i$ denotes the (fixed) neighbourhood of node i , the notation $[\leq t]$ indicates that the functions may access current and past representations up to time t , and $\rho_{\text{st}}^{(l)}, \gamma_{\text{st}}^{(l)}$ are learnable message and update functions that now incorporate the temporal state. Compared with the purely spatial message passing of (2.42), the STMP block must additionally integrate a sequence modelling mechanism to propagate information across time steps. In matrix form, the shorthand is

$$\mathbf{H}^{(l+1)}[t] = \text{STMP}^{(l)}(\mathbf{H}^{(l)}[\leq t], \mathcal{G}), \quad (2.50)$$

where $\mathbf{H}^{(l)}[\leq t]$ collects all node representations at layer l up to and including time t .

Most STGNN architectures share a common encode–process–decode framework [Cini, Marisca, Zambon, et al., 2025]. Given a window of graph signals, $\mathbf{X}[t-W+1], \dots, \mathbf{X}[t]$, where $\mathbf{X}[\tau] \in \mathbb{R}^{N \times d}$ and the shorthand $[a:b]$ denotes the index range, $a, a+1, \dots, b-1$, the model proceeds in three stages,

$$\mathbf{h}_i^{(0)}[\tau] = \text{Enc}(\mathbf{x}_i[\tau]), \quad \tau = t-W+1, \dots, t, \quad (2.51)$$

$$\mathbf{H}^{(l+1)}[t] = \text{STMP}^{(l)}(\mathbf{H}^{(l)}[\leq t], \mathcal{G}), \quad l = 0, \dots, L-1, \quad (2.52)$$

$$\hat{\mathbf{x}}_i[t+1:t+H] = \text{Dec}(\mathbf{h}_i^{(L)}[t]), \quad (2.53)$$

where $\text{Enc}(\cdot)$ and $\text{Dec}(\cdot)$ are pointwise encoding and decoding functions (e.g. linear layers or MLPs) that do not perform any spatiotemporal propagation, and H is the forecast horizon length. All spatiotemporal information exchange is delegated to the L stacked STMP layers in (2.52). Different STGNN architectures are therefore distinguished primarily by how the STMP block couples spatial aggregation with temporal modelling. The main design strategies can be categorised by

the order in which spatial and temporal operations are composed [Cini, Marisca, Zambon, et al., 2025].

Remark 12. *The encode–process–decode framework above is presented for the forecasting task, where the decoder maps the final representation, $\mathbf{h}_i^{(L)}[t]$, to future values, $\hat{\mathbf{x}}_i[t+1:t+H]$. The same architecture adapts to other spatiotemporal tasks by modifying the decoder and, where necessary, the temporal processing direction. For imputation, the model reconstructs missing values at observed time steps rather than forecasting ahead; this typically requires bidirectional temporal processing so that information from both past and future observations informs reconstruction at each time step [Cini, Marisca, and Alippi, 2022; Marisca, Cini, and Alippi, 2022]. For classification, the decoder maps representations to discrete class labels rather than continuous signals. In all cases, the STMP layers remain unchanged; the task is defined entirely by the choice of decoder, loss function, and masking strategy. Similarly, while the decoder in (2.53) produces point estimates, it can be replaced by a probabilistic output head that predicts distributional parameters (e.g. quantiles or mixture components [Benidis et al., 2022]). These task-specific adaptations are developed in Chapters 4, 5, 6, and 7.*

Time-and-Space Models

Time-and-space (T&S) models interleave spatial and temporal operations within each STMP block, so that each layer jointly conditions on graph structure and temporal history. The STMP update specialises (2.50) to

$$\mathbf{H}^{(l+1)}[t] = \text{STMP}^{(l)}(\mathbf{H}^{(l)}[t], \mathbf{H}^{(l+1)}[t-1], \mathcal{G}), \quad (2.54)$$

combining spatial aggregation on $\mathcal{G}$ with a temporal state update via $\mathbf{H}^{(l+1)}[t-1]$.

A commonly adopted instantiation, termed a graph recurrent neural network (GRNN), replaces the dense matrix multiplications inside a recurrent architecture, such as a GRU or LSTM gates, with graph message-passing operators. As a result, each gate aggregates spatiotemporal information from the neighbourhood of each node. For a GRU-based architecture at layer l [Seo et al., 2018; Yaguang Li et al., 2018], let $\mathbf{Z}^{(l)}[t] = \mathbf{H}^{(l-1)}[t]$ initialise the spatial input from the previous time

step. The gate and state updates are

$$\mathbf{R}^{(l)}[t] = \sigma\left(\text{MP}_r^{(l)}([\mathbf{Z}^{(l)}[t] \parallel \mathbf{H}^{(l)}[t-1]], \mathcal{G})\right), \quad (2.55)$$

$$\mathbf{U}^{(l)}[t] = \sigma\left(\text{MP}_u^{(l)}([\mathbf{Z}^{(l)}[t] \parallel \mathbf{H}^{(l)}[t-1]], \mathcal{G})\right), \quad (2.56)$$

$$\mathbf{C}^{(l)}[t] = \tanh\left(\text{MP}_c^{(l)}([\mathbf{Z}^{(l)}[t] \parallel \mathbf{R}^{(l)}[t] \odot \mathbf{H}^{(l)}[t-1]], \mathcal{G})\right), \quad (2.57)$$

$$\mathbf{H}^{(l)}[t] = \mathbf{U}^{(l)}[t] \odot \mathbf{H}^{(l)}[t-1] + (\mathbf{1} - \mathbf{U}^{(l)}[t]) \odot \mathbf{C}^{(l)}[t], \quad (2.58)$$

where $\parallel$ denotes concatenation, $\mathbf{R}^{(l)}[t]$ and $\mathbf{U}^{(l)}[t]$ are reset and update gates, $\mathbf{C}^{(l)}[t]$ is the candidate state, and $\text{MP}_r^{(l)}$, $\text{MP}_u^{(l)}$, $\text{MP}_c^{(l)}$ are message-passing operators (Fig. 2.2.5) with independent learnable parameters. The choice of operator determines the specific architecture: Yaguang Li et al. (2018) use diffusion convolutions based on powers of the random walk shift operator, Seo et al. (2018) employ spectral graph convolutions, and Zhang and Guo (2022) use graph attention within LSTM gates to capture heterogeneous diffusion patterns.

Beyond recurrent designs, T&S models can also be realised by integrating different temporal sequence models directly into the message function of the STMP block. For instance, B. Yu, H. Yin, and Zhu (2018) integrate a graph convolutional layer between two gated temporal convolutional layers, while Z. Wu et al. (2019) pair graph convolutions with dilated causal temporal convolutions. Cross-node temporal attention offers another mechanism for propagating information across sequences at neighbouring nodes [Marisca, Cini, and Alippi, 2022].

The time complexity of T&S models scales as $\mathcal{O}(W(N + L|\mathcal{E}|))$, since L message-passing layers are executed at each of the W input time steps. This dependence on the product of window length and the graph size (particularly the number of edges, $|\mathcal{E}|$) can become prohibitive for long sequences or dense graphs. However, the fine-grained coupling of spatial and temporal operations at each time step provides strong modelling capacity [Jenkins, Cini, Barker, Sharp, et al., 2025; Jenkins, Cini, Barker, D. Mandic, et al., 2025].

Time-then-Space Models

Time-then-space (TTS) models factorise the STMP block into two sequential stages: (i) a temporal encoder operating independently at each node, and (ii) spatial message passing over the encoded representations. Given the input window sequence, $\mathbf{X}[t-W+1:t]$, the temporal encoding step produces

$$\mathbf{h}_i^{(1)} = \text{SeqEnc}(\mathbf{x}_i[t-W+1:t]), \quad i = 1, \dots, N, \quad (2.59)$$

where $\text{SeqEnc}(\cdot)$ denotes a sequence encoder (e.g. an RNN, temporal convolutional network, or Transformer). The encoded representations, $\mathbf{H}^{(1)} = [\mathbf{h}_1^{(1)}, \dots, \mathbf{h}_N^{(1)}]^\top \in \mathbb{R}^{N \times d_1}$, are then propagated spatially through $L-1$ message-passing layers,

$$\mathbf{H}^{(l+1)} = \text{MP}^{(l)}(\mathbf{H}^{(l)}, \mathcal{G}), \quad l = 1, \dots, L-1. \quad (2.60)$$

Since temporal encoding operates independently per node at cost $\mathcal{O}(NW)$ and spatial message passing costs $\mathcal{O}(L|\mathcal{E}|)$, the total time complexity of a TTS model scales as $\mathcal{O}(NW + L|\mathcal{E}|)$. Note that, since the complexity is independent of W in the message-passing stage, the complexity scales favourably compared to T&S models, yielding substantial savings for long sequences or dense graphs [Cini, Marisca, Bianchi, et al., 2023; Gao and Ribeiro, 2022]. The trade-off for lower time complexity is a potential information bottleneck: the temporal encoder must compress the entire window into a fixed-size representation without access to spatial context from neighbouring nodes, which may be limiting, particularly when spatial and temporal dependencies are tightly coupled.

Reversing the factorisation order yields a space-then-time (STT) architecture, in which message-passing enriches node representations with spatial context at each time step before a final temporal encoder processes the resulting sequences. However, because message passing must execute at each of the W time steps, STT models incur the same $\mathcal{O}(W(N + L|\mathcal{E}|))$ time complexity as T&S models, but without the fine-grained spatiotemporal coupling that justifies this cost, making them uncommon in practice [Cini, Marisca, Zambon, et al., 2025; Gao and Ribeiro, 2022].

Remark 13. *The taxonomy above reflects a trade-off between representational power and scalability. T&S models maximise representational power by coupling spatial and temporal operations at every time step. TTS models gain scalability through factorisation, reducing time complexity to $\mathcal{O}(NW + L|\mathcal{E}|)$, but risking information bottlenecks. The classical graph VAR and CGP models of Sec. 2.3.1 can be viewed as linear instances of T&S models, whereby graph and temporal convolutions are jointly coupled through a time-vertex autoregression. Neural STGNN architectures generalise these principles through non-linear, learnable transformations, trading interpretability and sample efficiency for expressive power.*

2.3.3 Beyond Global Models, Fixed Topologies and Point Predictions

The graph-based spatiotemporal models presented in this chapter share parameters across all nodes and time steps, assume a known and fixed graph topology, $\mathcal{G}$, and produce deterministic point predictions. These assumptions enable parameter-efficient learning and principled generalisation, but each can become a limiting factor in practice. Relaxing them motivates the methods developed

in Chapters 5, 6, and 7.

Local effects in global architectures. The STGNNs presented in Sec. 2.3.2 share message-passing operators and temporal modules across all nodes, making them inherently global models: a single set of weights governs every node regardless of its position in the graph [Cini, Marisca, Zambon, et al., 2024]. This parameter sharing is a key strength, enabling both transductive inference at nodes seen during training and inductive generalisation to entirely new graphs with different sizes or topologies [Hamilton, Ying, and Leskovec, 2017]. However, global models assume that the dynamics at every node are generated by the same underlying spatiotemporal process. When this assumption is violated, for example because nodes exhibit heterogeneous local behaviour, a global model may require impractically large capacity or long input windows to implicitly distinguish between different regimes [Cini, Marisca, Zambon, et al., 2024]. In cardiac electrophysiology, the tissue substrate is inherently heterogeneous: spatially varying properties such as fibre orientation, conduction velocity, and fibrosis distribution (Sec. 3.2) govern the local propagation dynamics, while the electrophenotype spectrum (Sec. 3.3) implies that patients themselves occupy different positions along a continuum of dynamical complexity [F. S. Ng et al., 2020].

Hybrid global-local approaches address this by augmenting shared message-passing layers with a small number of node- or graph-specific learnable embedding parameters that condition the encoder and decoder on local context, effectively amortising the learning of specialised components without sacrificing the inductive transferability of the global architecture [Cini, Marisca, Zambon, et al., 2024]. Chapter 5 adopts this strategy, combining shared spatiotemporal message-passing layers with patient-specific and node-specific embeddings that capture unmeasured tissue properties, enabling personalised reconstruction of fibrillation dynamics across patients with diverse atrial substrates through rapid fine-tuning of only the embedding parameters.

Unknown and time-varying graph topologies. The graph topology, $\mathcal{G}$, and the graph shift operator, $\mathbf{S}$, have been treated throughout this chapter as given inputs to each model, from polynomial graph filters to STGNNs. Yet in many domains the relational structure is unknown and must be inferred from the observed signals. In neuroscience, functional connectivity between brain regions is estimated from co-activation patterns and may reconfigure during cognitive tasks [Preti, Bolton, and Van De Ville, 2017]; in finance, causal dependencies among assets shift during market regime changes [Papana et al., 2017]; and in cardiac fibrillation, preferential conduction pathways may underpin the fibrillation mechanism [Handa, X. Li, et al., 2020].

Batch approaches to graph topology estimation include sparse inverse covariance methods [Friedman, Hastie, and Tibshirani, 2007], signal-smoothness formulations [Xiaowen Dong, Thanou, et al., 2016], and causal graph learning methods that test whether the history of one signal improves prediction of another [Granger, 1969]. The latter framework extends naturally to graph-structured data through time-vertex autoregressive system identification, where the causal topology is parameterised as a graph shift operator within a graph VAR model [Mei and Moura, 2016]. However, all of these approaches estimate a single static topology from an entire dataset. Deep learning methods have also been developed to learn the graph structure end-to-end [Cini, Zambon, and Alippi, 2023; Shang and J. Chen, 2021], but their computational demands preclude real-time deployment on streaming data. When the topology is itself non-stationary, online algorithms that track changes in this structure with constant per-iteration cost become necessary. Chapter 7 develops such an algorithm, extending the CGP framework of Sec. 2.3.1 to the adaptive setting using the recursive estimation principles introduced in Sec. 2.1.4.

Uncertainty quantification for spatiotemporal predictions. The models in this chapter produce point estimates, but predictions deployed in safety-critical settings, from clinical decision-making to energy grid management, require quantified measures of reliability. Probabilistic spatiotemporal models can provide uncertainty estimates by parameterising predictive distributions, for example through quantile outputs [Benidis et al., 2022] or generative formulations [Rasul et al., 2021; Rühling Cachay et al., 2023]. However, these approaches impose distributional assumptions and are tightly coupled to the model architecture. Conformal prediction [Vovk, Gammerman, and Shafer, 2005; Angelopoulos, Bates, et al., 2023] offers an alternative distribution-free, post-hoc framework that wraps around any pre-trained predictor and constructs prediction intervals with finite-sample coverage guarantees derived from calibration residuals alone. The standard formulation assumes exchangeability between calibration and test data, an assumption violated by temporal dependence and spatial correlation [R. F. Barber et al., 2023]. Recent extensions handle temporal non-exchangeability through adaptive re-weighting of residuals [Gibbs and Candes, 2021; C. Xu and Y. Xie, 2023a; C. Xu and Y. Xie, 2023b; R. F. Barber et al., 2023; Auer et al., 2023], but operate independently on each time series without exploiting relational structure. Leveraging graph-based dependencies among correlated time series to sharpen conformal prediction intervals remains largely unexplored. Chapter 6 addresses this gap by introducing a conformal prediction method that uses graph deep learning operators to exploit relational structure in prediction residuals, providing calibrated uncertainty estimates for spatiotemporal forecasts.

Chapter 3

Cardiac Electrophysiology and Fibrillation

This chapter provides the clinical and electrophysiological foundations for the methods developed in this thesis. The chapter begins with the multi-scale nature of cardiac electrical activity, from ion channel dynamics at the cellular level to electrogram and ECG recordings at the organ and body-surface scales. Next, properties of tissue-level electrical propagation that govern wavefront dynamics are discussed, from conduction velocity to re-entry. Atrial fibrillation is introduced, reviewing the competing mechanistic hypotheses and the electrophenotype spectrum framework that unifies them, with direct implications for personalised treatment planning. Current treatment strategies are discussed, focusing on catheter ablation and the ceiling in long-term outcomes that arises from existing one-size-fits-all treatments. Finally, the clinical and research electrophysiological mapping modalities are surveyed, from invasive contact mapping to ex vivo optical mapping, highlighting the fundamental resolution–coverage trade-off that no single clinical mapping technology resolves, which motivates the data-driven, graph-based spatiotemporal reconstruction methods developed in this thesis.

3.1 Electrical Activity of the Heart	49
3.2 Electrical Propagation Dynamics	52
3.3 Mechanisms of Atrial Fibrillation	54
3.4 Treatment of Atrial Fibrillation	57
3.5 Electrophysiological Mapping	58
3.6 Summary and Research Questions	63

3.1 Electrical Activity of the Heart

The electrical activity of the heart spans multiple spatial scales, from the opening of ion channels in individual cells to the global electric field recorded on the body surface. Understanding what

signals arise at each scale provides essential context for the recordings that form the data for this thesis [Klabunde, 2011], though the methods developed herein operate at scales far above the single cell. Figure 3.1 illustrates this hierarchy.

Cellular scale. At rest, cardiomyocytes (heart muscle cells) maintain a membrane potential of approximately -90 mV, established by an unequal distribution of ions, principally sodium (Na^+), potassium (K^+), and calcium (Ca^{2+}), between the intracellular and extracellular spaces. The movement of these ions is governed by their electrochemical gradient, the net force arising from concentration differences and electrical potential across the membrane, and regulated by voltage-gated ion channels embedded in the cell membrane. When a cardiomyocyte receives sufficient stimulation from a neighbouring cell (above the activation threshold), voltage-gated sodium channels open rapidly, allowing an influx of Na^+ that drives the membrane potential sharply positive in a process known as depolarisation. A characteristic plateau phase follows, sustained by calcium influx and reduced potassium efflux, before repolarisation restores the resting potential as potassium channels reopen. This sequence of depolarisation, plateau, and repolarisation constitutes the cardiac action potential, lasting approximately 200–400 ms depending on cell type and heart rate [Klabunde, 2011].

Tissue scale. Cardiomyocytes are electrically coupled through gap junctions, protein channels that permit direct ion flow between adjacent cells. When one cell depolarises, the resulting current spreads to its neighbours, triggering their depolarisation in turn. This cell-to-cell conduction produces a propagating wavefront of activation that moves through cardiac tissue, with each cell firing after a short conduction delay from its predecessor [Klabunde, 2011]. The tissue-scale dynamics, including wavefront speed, direction, and morphology, are shaped by the properties discussed in Sec. 3.2 and are the spatiotemporal phenomena that the methods in this thesis aim to reconstruct and characterise.

Organ scale. A recording electrode positioned on the endocardial surface does not resolve individual cellular action potentials. Instead, it measures the cumulative electrical field produced by the synchronised depolarisation and repolarisation of thousands of neighbouring cardiomyocytes, yielding a voltage time series known as an (invasive) electrogram. Electrograms recorded from different positions across the heart capture the local passage of activation wavefronts, and their morphology reflects the underlying tissue properties and conduction patterns [Klabunde, 2011]. Invasive electrogram recordings, obtained during clinical electrophysiology procedures using multipolar catheters, form the primary data for the mapping methods developed in Chapter 5.

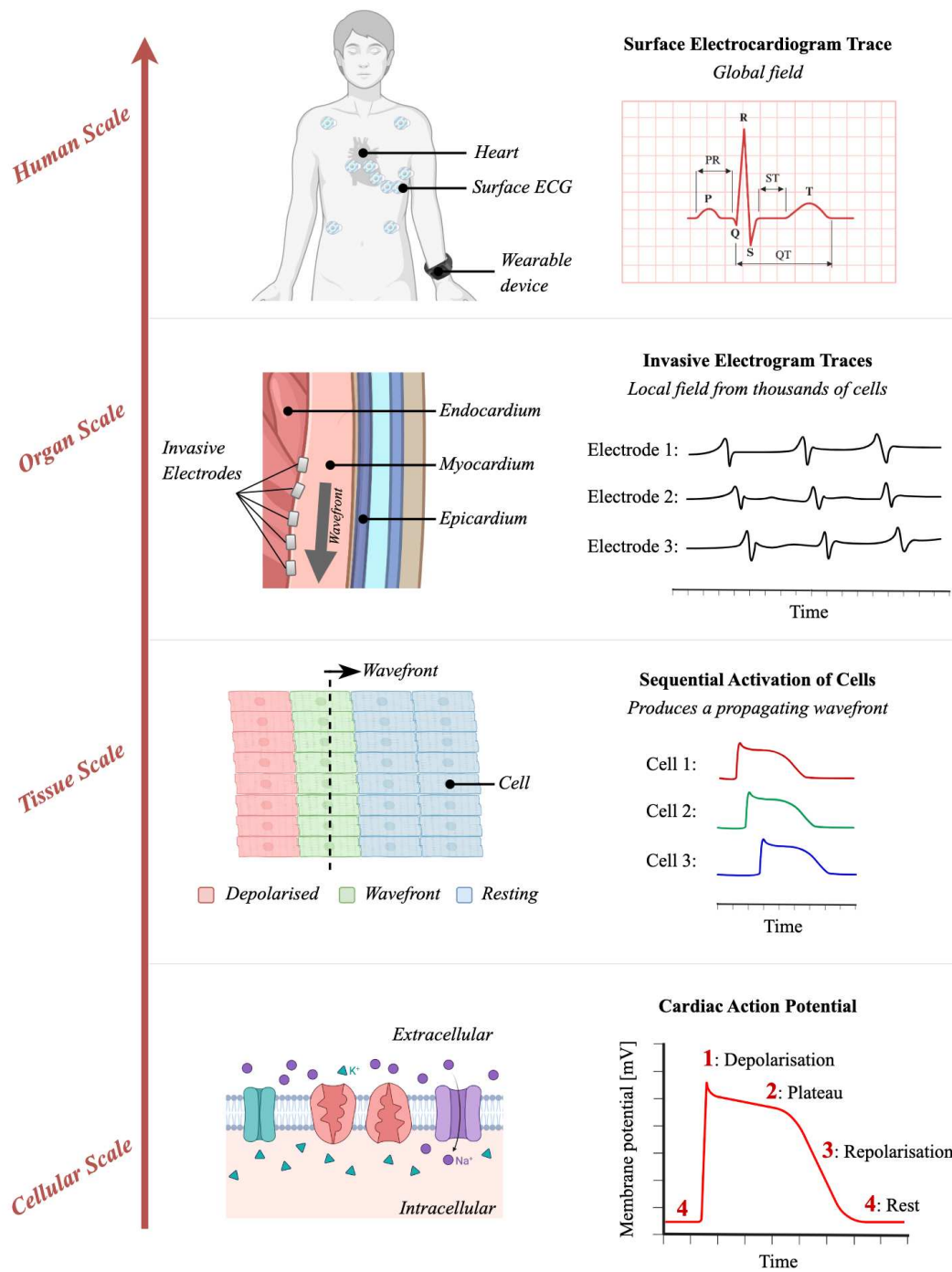


Figure 3.1: Cardiac electrical activity across spatial scales. At the cellular scale (bottom), voltage-gated ion channels regulate the flow of sodium (Na^+), potassium (K^+), and calcium (Ca^{2+}) ions across the cardiomyocyte membrane, producing the action potential with its phases of depolarisation (1), plateau (2), repolarisation (3), and rest (4). At the tissue scale (middle-bottom), sequential depolarisation of electrically coupled cells produces a propagating wavefront of activation. At the organ scale (middle-top), invasive electrodes on the endocardial surface record electrograms reflecting the cumulative field of thousands of cells. At the human scale (top), surface electrodes capture the ECG, the global electrical signature of the heart, recorded either in the clinic or via wearable devices. The signal processing and machine learning methods in this thesis operate at the organ and human scales.

Human scale. At the body surface, the aggregate electrical activity of the entire heart produces the ECG, a non-invasive recording of the heart’s global electric field. The standard 12-lead ECG, recorded from electrodes placed on the chest and limbs, captures characteristic waveform features, including the P wave (atrial depolarisation), QRS complex (ventricular depolarisation), and T wave (ventricular repolarisation), that reflect the sequential activation and recovery of cardiac chambers [Klabunde, 2011]. Increasingly, single- or few-lead ECG recordings from wearable devices provide continuous cardiac monitoring, enabling the detection of arrhythmias such as atrial fibrillation outside the clinical setting [Steinhubl et al., 2018]. The temporal imputation methods developed in Chapter 4 operate on ECG signals at this scale.

3.2 Electrical Propagation Dynamics

At the tissue level, the physics of electrical propagation is formally described by reaction-diffusion partial differential equations. The standard formulation is the monodomain equation, which models current flow through cardiomyocytes connected via gap junctions, coupled with a system of ordinary differential equations representing the nonlinear dynamics of ion channels, pumps, and exchangers at each cell [Pullan, L. K. Cheng, and Buist, 2005]. When extracellular current flow must be accounted for explicitly, the more general bidomain formulation is used [Tung, 1978; Trayanova and Winslow, 2011]. While this reaction-diffusion framework is well established, its practical application requires specifying a cellular ionic model, of which many exist at different levels of biophysical detail [Noble and Rudy, 2001], together with patient-specific parameters including tissue conductivities, fibre orientation, and the extent of structural remodelling such as fibrosis [Trayanova and Winslow, 2011; Richard A. Gray and Pathmanathan, 2018]. These quantities vary between patients and are difficult to obtain *in vivo*, limiting the clinical applicability of purely simulation-based approaches [Trayanova and Winslow, 2011]. The methods developed in this thesis are data-driven and do not rely on these biophysical models directly; however, several macroscopic properties of cardiac propagation that emerge from this physics are central to the signal processing challenges addressed in later chapters.

Conduction velocity. The speed at which the activation wavefront propagates through cardiac tissue is known as the conduction velocity. This is typically 0.5–1.0 m/s in healthy atrial tissue [Kleber and Rudy, 2004]. Conduction velocity is not spatially uniform: it depends on local tissue properties including cell-to-cell coupling strength, fibre orientation, and the presence of fibrosis or scar tissue. Regions of dense fibrosis, for example, force wavefronts to navigate meandering paths around insulating collagen deposits, producing local conduction slowing and

increasing the likelihood of wavebreak [Nguyen, Qu, and Weiss, 2014]. This spatial heterogeneity means that wavefronts do not propagate as simple expanding wavefronts but instead follow complex, patient-specific pathways determined by the underlying tissue substrate.

Anisotropy. Cardiac muscle fibres are organised in a highly structured arrangement, with conduction velocity along the fibre direction typically two to three times faster than in the transverse direction [Kleber and Rudy, 2004]. This anisotropy arises because gap junctions are predominantly located at the ends of the elongated cardiomyocytes rather than along their lateral surfaces, facilitating faster longitudinal than transverse current flow. In the atria, fibre architecture is particularly complex, with abrupt changes in fibre orientation across thin-walled regions and between the endocardial and epicardial layers [S. Y. Ho, Anderson, and Sánchez-Quintana, 2002]. This structural complexity contributes to the spatially heterogeneous propagation patterns observed during both normal rhythm and fibrillation, and is one reason why functional connectivity between regions of tissue may differ substantially from what simple anatomical proximity would suggest.

Refractoriness. Following depolarisation, a cardiomyocyte enters a refractory period during which it cannot be re-excited regardless of the strength of the stimulus. This property arises from the inactivation of sodium channels during repolarisation and ensures that the activation wavefront propagates forward without immediately re-exciting tissue it has already moved through. The duration of the refractory period, known as the effective refractory period, varies across the atrial surface and is modulated by factors including heart rate, autonomic tone, and disease [Janse, 1997; Coumel, 1994]. Spatial heterogeneity in refractoriness has profound consequences for arrhythmia susceptibility, as discussed in the following subsection.

Re-entry and wave break. In normal sinus rhythm, the electrical wavefront initiated by the sinoatrial node propagates in an orderly fashion across the atria, producing coordinated contraction. Under pathological conditions such as atrial fibrillation, however, the interplay between heterogeneous conduction velocity, anisotropic fibre structure, and non-uniform refractoriness can give rise to far more complex activation patterns.

The concept of wavelength, which is defined as the product of conduction velocity, CV, and the effective refractory period (ERP), defined as $\lambda = CV \times ERP$, represents the spatial extent of the refractory tail behind a propagating wavefront and provides a useful framework for understanding when re-entry can occur [Kleber and Rudy, 2004; Rensma et al., 1988]. Re-entry requires that a wavefront travel a path long enough for the tissue at its origin to become re-excitable, which is only possible when the circuit length exceeds the wavelength; leaving an excitable gap of recovered

tissue ahead of the front. These re-entrant circuits generally manifest in two forms:

- *Anatomical re-entry*: The impulse circulates around a fixed, non-excitable physical obstacle (such as scar or a vessel). In this case, the circuit path is determined by the anatomy, and the excitable gap is often large [Kleber and Rudy, 2004].
- *Functional re-entry*: The impulse circulates without a physical obstacle, instead pivoting around a core of refractory tissue. This takes the form of rotating spiral waves (rotors) that pivot around a central phase singularity, the point/core of the spiral wave where the phases of the action potential converge [Richard A Gray, Pertsov, and Jalife, 1998; Winfree, 1994]. Unlike anatomic circuits, functional re-entry is highly dynamic, the point/core can move, and the circuit length is often nearly equal to the wavelength, resulting in a very small or non-existent excitable gap.

When the dispersion of refractoriness is sufficiently high, a single spiral wave can fragment via localised wave break into multiple independent wavelets, each generating a complex, stochastic spatiotemporal pattern [Weiss et al., 2005; F. H. Fenton et al., 2002]. It is these phenomena (spiral waves, phase singularities, and wavebreak), that underlie the mechanisms of cardiac fibrillation discussed in Sec. 3.3, and the spatiotemporal complexity they produce is what makes fibrillation so challenging to map from sparse electrode recordings.

3.3 Mechanisms of Atrial Fibrillation

Cardiac fibrillation is the breakdown of the orderly wavefront propagation described in the previous section into rapid, irregular, and spatiotemporally complex electrical activity. In AF, this disorganisation occurs in the atrial (upper) chambers of the heart, producing uncoordinated atrial contraction and an irregular ventricular rate. AF is broadly classified as paroxysmal (terminates spontaneously or with intervention within 7 days of onset), persistent (continuously sustained beyond 7 days), or long-standing persistent (continuous for more than 12 months) [Hindricks et al., 2021]. Disease progression often follows the “AF begets AF” paradigm [Wijffels et al., 1995], where the arrhythmia induces electrical and structural remodelling, such as increased fibrosis, that shifts the focus from initial triggers to a self-sustaining substrate [Heijman et al., 2014]. Figure 3.2 contrasts the organised electrical conduction of normal sinus rhythm with the disorganised activation patterns characteristic of AF.

Despite decades of investigation, the electrophysiological mechanisms that sustain AF remain disputed, with no definitive consensus [F. S. Ng et al., 2020]. Historically, the debate has been defined

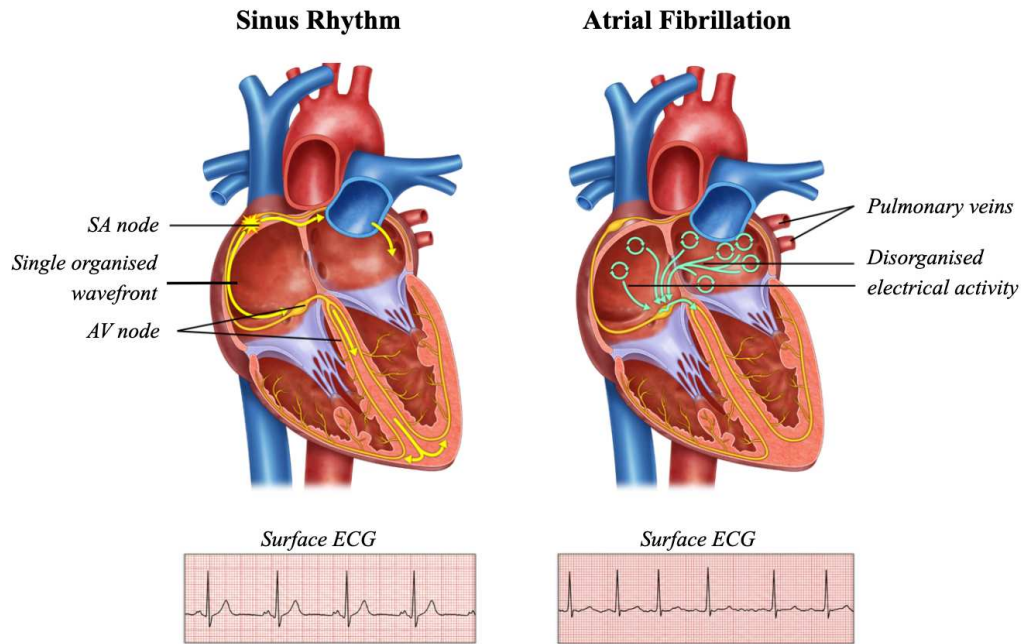


Figure 3.2: Comparison of normal sinus rhythm and atrial fibrillation. Left: During sinus rhythm, the sinoatrial (SA) node initiates a single organised wavefront that propagates through the atria to the atrioventricular (AV) node, producing regular P waves on the surface ECG. Right: During atrial fibrillation, disorganised electrical activity (often triggered by ectopic foci near the pulmonary veins [Haissaguerre et al., 1998]) replaces the coordinated wavefront with multiple chaotic wavelets, resulting in irregular atrial activation and an irregular ventricular response on the surface ECG, with absence of discrete P waves. Adapted from [Melbourne Heart](#).

by two divergent frameworks. The *multiple-wavelet hypothesis*, established through the computational modelling of Moe, Rheinboldt, and Abildskov (1964) and later experimental validation by Allessie et al. (2010) and Eckstein et al. (2013), describes an “anarchical model” where chaos is sustained by the continuous collision and regeneration of independent wavelets. Conversely, the *focal driver hypothesis* or “hierarchical model” suggests that AF is governed by discrete, localised sources, such as spiral waves, which drive the surrounding myocardium into fibrillatory conduction [Pandit and Jalife, 2013]. While this latter model gained significant traction following the clinical mapping studies of Narayan, Krummen, et al. (2012), it remains a subject of intense scrutiny. Recent high-resolution contact mapping has frequently failed to validate the presence of stable rotational activity in human AF, suggesting that earlier observations of localised drivers may have been artifacts of lower-resolution recording techniques [G. Lee et al., 2014; S. Lee et al., 2017; Roney et al., 2017].

A potentially unifying framework to resolve this dichotomy is the *electrophenotype spectrum* proposed by F. S. Ng et al. (2020), which suggests that these mechanisms are not mutually exclusive but represent the extremes of a functional continuum (Figure 3.3). This model proposes that the

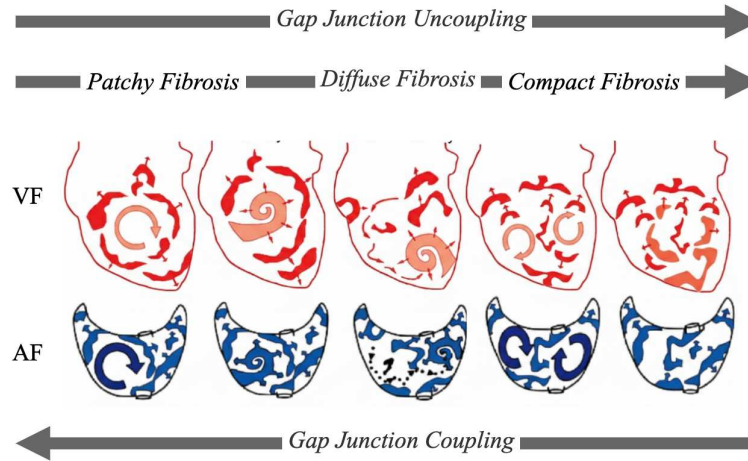


Figure 3.3: The electrophenotype spectrum of cardiac fibrillation, illustrating how progressive structural remodelling, from gap junction uncoupling (top) through patchy, diffuse, and compact fibrosis, gives rise to a continuum of fibrillatory mechanisms in both ventricular fibrillation (VF, red) and atrial fibrillation (AF, blue). At one extreme (left), well-coupled tissue with limited fibrosis supports organised activity sustained by stable rotational drivers. At the other (right), extensive structural disruption produces disorganised multi-wavelet re-entry. Intermediate substrates give rise to a range of spatiotemporal patterns, including meandering rotational dynamics and mixed driver-wavelet dynamics. This framework implies that effective mechanism-directed therapy requires first identifying where an individual patient lies on this spectrum, then selecting an appropriate intervention: targeted driver ablation for organised phenotypes versus atrial compartmentalisation or pharmacological approaches for disorganised phenotypes [F. S. Ng et al., 2020]. Adapted from F. S. Ng et al. (2020).

specific manifestation of fibrillation is an emergent property of the underlying atrial substrate. Studies in ventricular fibrillation (VF) models by Handa, X. Li, et al. (2021) have demonstrated that the spatial distribution of fibrosis and the degree of gap junction coupling determine positioning along this spectrum. Specifically, preserved gap junction coupling and patchy fibrotic patterns tend to anchor stable rotational drivers, manifesting as an organised electrophenotype. Conversely, compact fibrosis or significant gap junction uncoupling promotes the breakdown of wavefronts into the disorganised, multi-wavelet dynamics characteristic of the anarchical model. F. S. Ng et al. (2020) and Handa, X. Li, et al. (2020) hypothesise a parallel spectrum for AF, where the cumulative degree of remodelling determines the predominant, patient-specific mechanism. Consequently, AF is viewed as a dynamic state that shifts along this continuum as the underlying substrate progressively deteriorates.

The electrophenotype framework has important implications for treatment. If the mechanism sustaining AF varies between patients, then an effective mechanism-directed treatment could first identify the patient's electrophenotype and tailor the therapeutic strategy accordingly toward it [F. S. Ng et al., 2020]. This vision is currently limited by the inability to accurately map AF from the sparse data available during clinical procedures, as discussed in the following

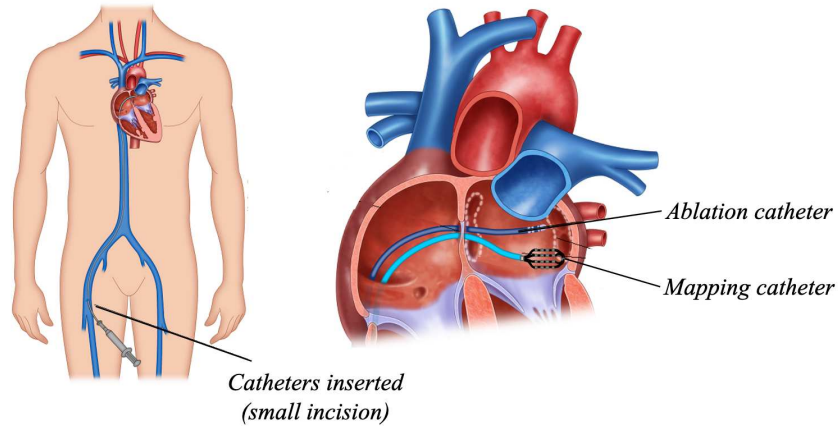


Figure 3.4: Overview of catheter ablation for atrial fibrillation. Left: Catheters are introduced via a small incision in the groin and advanced through the vasculature to the heart. Right: Within the left atrium, a multipolar mapping catheter records local electrical activity (electrograms) to characterise activation patterns, while a separate ablation catheter delivers targeted energy to create lesions that electrically isolate arrhythmogenic tissue. The sparse spatial coverage of the mapping catheter relative to the atrial surface area is a central limitation addressed in this thesis. Adapted from [Melbourne Heart](#).

sections. From a signal processing perspective, the electrophenotype spectrum also underscores a key challenge for the methods developed in this thesis: the spatiotemporal complexity of fibrillation varies substantially between patients and may change over time within a single patient, demanding reconstruction and analysis methods that can accommodate a wide range of dynamical behaviours. The functional connectivity patterns learned by ADACGP (Chapter 7) provide a data-driven characterisation of propagation patterns from fibrillation recordings, complementing the spatial reconstruction provided by FIBMAP (Chapter 5); integrating these components into a unified reconstruct-then-characterise pipeline is a direction for future work (Chapter 8).

3.4 Treatment of Atrial Fibrillation

The clinical management of AF broadly comprises two complementary strategies: rate control, which accepts the presence of AF but limits the ventricular response using pharmacological agents; and rhythm control, which aims instead to restore and maintain sinus rhythm using anti-arrhythmic drugs and/or interventional procedures [Hindricks et al., 2021]. Anti-arrhythmic drugs can suppress AF recurrences and reduce symptom burden, but are constrained by modest efficacy, organ-specific toxicities, and the requirement for long-term administration [Hindricks et al., 2021]. For symptomatic patients in whom drug therapy is insufficient or poorly tolerated, catheter ablation (Figure 3.4) has become the dominant rhythm-control strategy [Cardoso et al., 2022; L.-Z. Shi et al., 2015; Piccini and Fauchier, 2016]. However, despite advances in catheter technology and ab-

lation delivery, treatment remains fundamentally constrained by its empirical nature: standardised lesion sets are applied regardless of the underlying substrate, contributing to a persistent ceiling in long-term success rates. This is particularly true in persistent AF, where success rates are as low as 50% at five years [Scherr et al., 2015; Verma et al., 2015].

The cornerstone of AF ablation is pulmonary vein isolation, which targets the ectopic triggers frequently originating within the pulmonary veins [Haissaguerre et al., 1998]. While effective in paroxysmal AF where structural remodelling is minimal, pulmonary vein isolation alone is frequently insufficient for persistent cases, in which a self-sustaining arrhythmogenic substrate can maintain fibrillation independently of the original triggers [Heijman et al., 2014; Nattel, Burstein, and Dobrev, 2008]. Efforts to modify this substrate have targeted various electrophysiological and structural properties, including linear lesions to compartmentalise atrial tissue [Jais et al., 2004], ablation of complex fractionated atrial electrograms at presumed driver sites [Nademanee et al., 2004], and localisation of rotational activity or fibrotic regions [Narayan, Krummen, et al., 2012; A. S. Jadidi et al., 2016]. However, large-scale trials, most notably “STAR-AF 2”, failed to demonstrate clear benefit from these adjunctive strategies, prompting a more pragmatic emphasis on durable pulmonary vein isolation as the primary evidence-based intervention [Verma et al., 2015; Mohanty et al., 2018].

The neutral outcomes of these trials underscore a fundamental limitation: empirical targeting rules are applied without first identifying the specific mechanism sustaining AF in a given patient. As established by the electrophenotype spectrum framework (Figure 3.3), a patient at the organised end of the continuum likely requires a fundamentally different therapeutic approach than one exhibiting disorganised, multi-wavelet dynamics [F. S. Ng et al., 2020]. Transitioning from empirical to mechanism-directed therapy is therefore contingent on the ability to robustly resolve these phenotypes from clinical data, a challenge that motivates the mapping and signal processing methods developed in this thesis.

3.5 Electrophysiological Mapping

The treatment strategies discussed in the previous section are fundamentally constrained by the clinician’s ability to observe electrical activity across the cardiac surface. Electrophysiological mapping refers to the acquisition and spatial registration of electrical recordings from the heart, with the aims of characterising the mechanism and propagation dynamics of arrhythmias and identifying suitable targets for treatment [Sy, Thiagalingam, and Stiles, 2012]. A central tension exists behind every electrophysiological mapping modality: no single technology simultaneously deliv-

ers high spatial resolution, temporal simultaneity, full chamber coverage, and high signal fidelity. Each modality resolves this tension through a different engineering trade-off, and the resulting data limitations define the signal processing challenges that the graph-based reconstruction methods of this thesis are designed to address.

This section focuses on the modalities whose data are used in this thesis: invasive contact and non-contact mapping, which provide the clinical recordings, and optical mapping, which provides the high-resolution ground truth for model development and validation. Additional modalities exist in practice, including non-invasive electrocardiographic imaging [Rudy and Burnes, 1999] and epicardial contact mapping during open-chest surgery [De Groot et al., 2022], but these are largely confined to the research setting and are omitted here to maintain focus. Table 3.1 summarises the key characteristics of the modalities discussed.

3.5.1 Clinical Invasive Mapping

Invasive mapping during electrophysiology procedures encompasses two broad paradigms: local contact mapping, which sequentially acquires high-resolution recordings from localised tissue regions using a roving multipolar catheter, and global mapping, which captures electrical activity from the entire atrial chamber simultaneously [De Groot et al., 2022]. These paradigms, illustrated in Figure 3.5, represent complementary positions on the resolution-coverage spectrum, and their respective limitations motivate the spatiotemporal graph reconstruction framework developed in Chapter 5.

Local Mapping

In local contact mapping, a multipolar catheter is placed against the endocardial surface and sequentially repositioned to acquire electrograms across the chamber [De Groot et al., 2022].

Table 3.1: Comparison of electrophysiological mapping modalities used in this thesis. Each modality resolves the resolution–coverage–simultaneity tension through a different engineering trade-off.

Modality	Spatial Coverage	Temporal Nature	Primary Trade-Off
Contact catheter	Sparse, local	Sequential	High near-field fidelity at the cost of non-contemporaneous acquisition
Global non-contact	Full chamber	Simultaneous	Whole-chamber view at the cost of low spatial resolution (due to inverse reconstruction)
Optical (research)	Full field	Simultaneous	Dense ground truth at the cost of <i>ex vivo</i> preparation

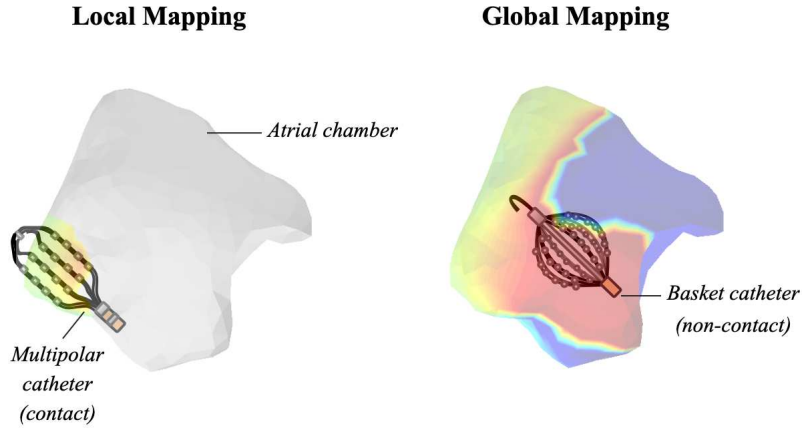


Figure 3.5: The two paradigms of clinical invasive mapping. Left: Local contact mapping, in which a multipolar catheter records high-fidelity electrograms from a small region of the atrial endocardium. The catheter must be sequentially repositioned to cover the chamber, yielding sparse, non-contemporaneous observations. Right: Global non-contact mapping, in which a basket catheter positioned within the atrial cavity records intra-cavitary potentials that are projected onto the entire endocardial surface via an inverse solution, providing a simultaneous whole-chamber view at the cost of spatial smoothing inherent in the regularised reconstruction.

Recordings may be obtained in unipolar mode, where the potential at a tissue-contact electrode is measured relative to a distant reference, providing an unambiguous local activation time at the cost of susceptibility to far-field contamination; or in bipolar mode, where the potential difference between adjacent electrodes suppresses common far-field signals but introduces dependence on electrode spacing, local conduction velocity, and wavefront orientation [De Groot et al., 2022]. In either configuration, the direct tissue contact yields higher near-field signal fidelity than non-contact or non-invasive modalities [Sy, Thiagalingam, and Stiles, 2012].

High-density multipolar catheters such as the Advisor™ HD Grid (16 electrodes, 4×4 arrangement, 3 mm spacing) provide substantially finer local detail than basket catheters, with the planar geometry promoting more consistent electrode-tissue contact and more uniform spatial sampling [De Groot et al., 2022; Narayan and John, 2024]. The patient-specific chamber geometry on which signals are processed is constructed progressively as the catheter is swept across the endocardium: each contact point contributes both a spatial coordinate and a local electrogram, yielding a three-dimensional surface mesh, typically containing 10^3 to 10^4 vertices [Sy, Thiagalingam, and Stiles, 2012; Narayan and John, 2024]. Modern electroanatomical mapping platforms (CARTO, EnSite, Rhythmia) localise the catheter using magnetic, impedance-based, or hybrid tracking with bench-tested positional accuracies on the order of 1 mm, and can incorporate pre-procedural CT or MRI surface models through image registration [Sy, Thiagalingam, and Stiles, 2012; Narayan and John, 2024].

For stable arrhythmias such as focal or macro-re-entrant atrial tachycardias, where the activation pattern is effectively constant from beat to beat, local activation times from sequential catheter positions can be reliably assembled into a global activation map that accurately localises the arrhythmic source [Özgül et al., 2023]. This approach breaks down during AF, where the disorganised activation pattern varies on every beat, meaning that recordings from different spatial positions and different time instants cannot be straightforwardly combined. Reconstructing the global spatiotemporal dynamics of AF from these sparse, asynchronous observations is the central challenge addressed by FIBMAP in Chapter 5.

Global Mapping

The limited instantaneous spatial coverage of contact mapping has motivated the development of global mapping approaches that simultaneously record electrical activity from the entire atrial chamber [De Groot et al., 2022]. These systems achieve their panoramic view by trading local signal fidelity for spatial completeness, a trade-off that is nonetheless valuable for AF, whose beat-to-beat variability precludes the temporal alignment that sequential contact mapping requires.

Multi-electrode basket catheters provide simultaneous low-resolution contact recordings by distributing flexible splines carrying up to 64 electrodes around the endocardial cavity. In the focal impulse and rotor modulation (FIRM) workflow, phase-based algorithms applied to these recordings identified putative rotor cores and focal drivers, with early single-centre studies reporting encouraging ablation outcomes [Narayan, Krummen, et al., 2012]. Subsequent multicentre trials, however, produced heterogeneous results [Brachmann et al., 2019; Mohanty et al., 2018]. The inconsistencies can be attributed in part to non-uniform spline spacing and poor electrode-tissue contact, with only 63% of electrode pairs on the 64-electrode basket falling within the 14.2 mm spacing required for reliable detection of rotational activity [Roney et al., 2017], and structures such as the septum and coronary sinus fall outside the basket’s physical reach [De Groot et al., 2022].

Non-contact global mapping systems such as AcQMap (Acutus Medical, Inc.) begin to address these limitations by measuring intra-cavitary far-field potentials from the centre of the chamber without requiring direct tissue contact [Grace et al., 2019]. The AcQMap catheter combines 48 ultrasound transducers with 48 electrodes on a spheroidal array: the ultrasound elements rapidly acquire a dense point cloud of the endocardial surface, while the electrodes record intra-cavitary potentials. A forward propagation model relates hypothetical endocardial charge densities to the measured potentials, and this relationship is inverted with regularisation to yield estimated charge-density distributions from which global activation maps are computed [Grace et al., 2019].

Non-contact reconstruction is an ill-posed inverse problem in which a small set of electrode measurements must be used to estimate a far larger set of surface quantities. Accuracy deteriorates with increasing array-to-wall distance, becoming unreliable beyond approximately 40 mm [Thiagalingam et al., 2004]. The combination of undersampling and regularisation-induced smoothing can generate apparent rotational activity with no counterpart in the true activation field [Roney et al., 2017; King et al., 2017], and independent verification using high-density contact catheters has often failed to confirm the stable rotor patterns reported by non-contact and phase-based mapping studies [G. Lee et al., 2014; S. Lee et al., 2017]. As a result, global mapping systems are yet to achieve guideline-level endorsement for routine clinical use beyond pulmonary vein isolation [De Groot et al., 2022].

Despite these limitations, non-contact mapping data provide the largest clinical dataset used in this thesis. The simultaneous whole-chamber recordings from 51 patients with human AF, comprising over 4×10^9 unique spatiotemporal observations, serve as the primary clinical data source for FIBMAP (Chapter 5). The non-contact reconstructions are treated as a clinically available and feasible approximation, against which the spatiotemporal graph reconstruction from sparse simulated catheter placements is evaluated; accordingly, all reconstruction performance metrics should be interpreted in the context of the inverse-problem limitations of the source data. The virtual electrogram positions on the endocardial surface mesh define the graph nodes, with edges determined by spatial adjacency on the reconstructed geometry.

3.5.2 Optical Mapping

While the clinical modalities described above provide the patient data used in this thesis, the high-resolution simultaneous ground truth required for model development and validation comes from optical mapping, a research modality that is unavailable during clinical procedures, yet defines the mechanistic reference standard for studying fibrillation dynamics [Efimov, Nikolski, and Salama, 2004]. Excised cardiac tissue sustained by Langendorff perfusion is stained with a voltage-sensitive fluorescent dye and illuminated at the dye’s excitation wavelength. Membrane depolarisation shifts the dye’s emission spectrum, producing fluorescence changes proportional to the local transmembrane potential. A high-speed camera (typically 1 to 4 kHz) images these fluorescence transients across the exposed tissue surface, yielding a spatiotemporally dense recording in which every pixel functions as an independent virtual electrode [Efimov, Nikolski, and Salama, 2004]. Figure 3.6 illustrates this experimental setup.

The pixel-wise nature of optical recordings supports a rich set of derived analyses that have shaped the mechanistic understanding of fibrillation [Davidenko et al., 1992; Richard A Gray, Pertsov,

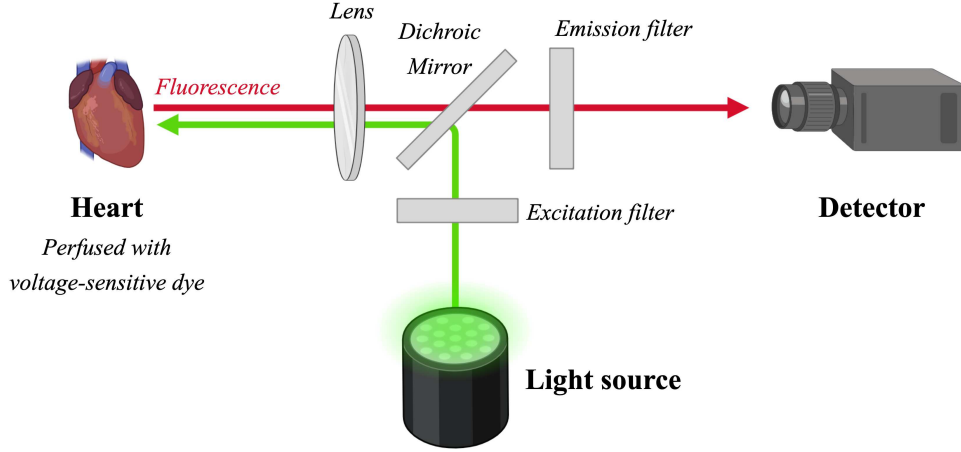


Figure 3.6: Optical mapping of cardiac fibrillation. Voltage-sensitive fluorescent dyes are applied to excised cardiac tissue, and high-speed cameras record the fluorescence changes accompanying action potential propagation across the tissue surface. Each pixel provides an independent time series of the local transmembrane potential, yielding a spatiotemporally dense ground-truth recording for the reconstruction and graph learning methods developed in Chapters 5 and 7, respectively.

and Jalife, 1998]. However, several constraints must be noted. The preparation is *ex vivo*: contractile motion must be suppressed with agents such as Blebbistatin for accurate imaging, which itself alters action potential duration [Kappadan et al., 2020]. Fluorescence signals arise predominantly from superficial tissue layers, so transmural activation patterns and three-dimensional scroll-wave dynamics cannot be directly observed [De Groot et al., 2022]. As a result, the mechanistic phenomena established by optical mapping (rotational activity, dominant-frequency gradients, phase singularities) can only be identified in clinical practice through invasive electrogram recordings [De Groot et al., 2022].

For the graph-based methods of this thesis, the pixel locations are treated as graph nodes on a two-dimensional regular grid, with edges connecting spatially adjacent pixels. The regular grid structure and simultaneous full-field recording make optical mapping data well suited for validating spatiotemporal reconstruction: the ground truth is available at every spatial location simultaneously. Optical mapping serves as the primary validation dataset for FIBMAP (Chapter 5) and for demonstrating ADACGP (Chapter 7), where the fully-observed signals are required for graph topology learning.

3.6 Summary and Research Questions

The preceding survey reveals that no existing mapping modality simultaneously delivers the spatial resolution, temporal simultaneity, and full-chamber coverage required to characterise the complex

spatiotemporal dynamics of AF in clinical practice. Contact mapping catheters deliver the highest signal fidelity but covers only a fraction of the atrial surface at any instant, and its sequential nature is incompatible with the beat-to-beat variability of AF. Global mapping systems achieve simultaneous whole-chamber coverage but rely on ill-posed inverse solutions whose regularisation limits spatial resolution and can introduce artefactual features. Research modalities such as optical mapping provide dense, simultaneous ground truth but are confined to *ex vivo* settings. These limitations, combined with the electrophenotype spectrum’s implication that effective therapy requires patient-specific mechanistic characterisation, collectively motivate the three research questions of this thesis.

RQ1 addresses the most immediate bottleneck: reconstructing global AF dynamics from the sparse local mapping observations available during routine clinical procedures. Rather than relying on inverse solvers, FIBMAP (Chapter 5) learns this mapping directly from high-resolution training data, operating on the patient-specific atrial geometry represented as a graph. **RQ2** recognises that clinical decisions require not only point predictions but also calibrated prediction uncertainty; COREL (Chapter 6) provides this through conformal prediction intervals for spatiotemporal forecasts. **RQ3** moves beyond reconstruction to characterise fibrillation mechanisms; ADACGP (Chapter 7) learns the time-varying functional connectivity patterns that reflect the underlying propagation dynamics, offering an objective, data-driven approach to identifying the fibrillation mechanism from dense spatiotemporal observations. FIBMAP directly addresses the inability to reconstruct global AF dynamics from sparse clinical observations, while COREL and ADACGP develop the methodological foundations needed to realise a robust, uncertainty-aware mechanism identification pipeline. Collectively, these contributions already enable capabilities that no existing mapping modality provides: FIBMAP reconstructs global AF dynamics from sparse routine catheter measurements validated on 51 patients; ADACGP characterises dynamic propagation patterns from fibrillation recordings where batch methods fail. While full integration of these components into a unified clinical pipeline remains a direction for future work (Chapter 8), each method independently addresses a concrete limitation of current clinical practice.

Part II

Imputation and Reconstruction

Chapter 4

Improving Diffusion Models for ECG Imputation

This chapter addresses the imputation of missing values in ECG signals recorded by wearable devices, where poor electrode contact and connectivity issues introduce artefacts and gaps that degrade downstream clinical analysis. PULSEDIFF, the template-guided denoising diffusion probabilistic model proposed in this chapter, exploits the quasi-periodic structure of the ECG through three innovations: extraction of a subject-level waveform template from observed data, beat-level stochastic augmentation to capture morphological variability across cardiac cycles, and a rhythm confidence score that adapts prior strength based on rhythmic regularity. Validated on the PTB-XL dataset, PULSEDIFF improves the reconstruction accuracy of two baseline diffusion models, outperforms the leading deterministic method for transient missing data, and remains competitive for extended gaps. Beyond its standalone contributions, this chapter demonstrates that temporal priors can substantially improve imputation of quasi-periodic signals while highlighting their insufficiency for the chaotic dynamics of cardiac fibrillation, motivating the graph-based spatial representations developed in the subsequent chapter.

4.1	Introduction	67
4.2	Preliminaries	70
4.3	Methodology	73
4.4	Experiments	80
4.5	Discussion	87
4.6	Summary and Relation to Thesis	89

This chapter is based on Jenkins, Z. Chen, et al. (2023).

4.1 Introduction

The spatiotemporal methods of [Chapter 2](#) exploit both relational and temporal structure to model signals on graphs. Before addressing the full spatiotemporal reconstruction problem directly, this chapter explores a foundational question: how effectively can temporal priors alone guide the reconstruction of missing physiological signals? The ECG, recorded at the body-surface scale ([Sec. 3.1](#)), provides an ideal setting in which to investigate this question. Its quasi-periodic structure, comprising repeating cardiac cycles with characteristic P-wave, QRS-complex, and T-wave morphologies, offers strong temporal priors that can be exploited by generative models. This chapter introduces PULSEDIFF, a template-guided denoising diffusion probabilistic model for ECG imputation across both healthy and pathological recordings. PULSEDIFF leverages quasi-periodic temporal priors while highlighting the fundamental limitations of purely temporal approaches for signals whose dynamics lack quasi-periodic structure, such as the chaotic electrical activity of cardiac fibrillation.

Advances in wearable technology have enabled passive real-time monitoring of personal health through recordings of physiological signals in mobile health (mHealth) systems. Among these, quasi-periodic physiological recordings, collectively termed pulsative signals, such as the ECG [Harry J Davies et al., [2024](#); M. Xu et al., [2022](#)], photoplethysmography [Harry J. Davies et al., [2022](#)], and blood pressure waveforms [Z. Chen, Dees, and Danilo P. Mandic, [2020](#)] are extensively collected because of their close relationship with cardiovascular and circulatory health. However, in comparison with clinical monitoring systems, a major challenge in mHealth is the prevalence of missing and corrupted data, arising for example from poor electrode contact during the recording of the ECG. Once identified, corrupted segments can be excised and treated as missing values, framing artefact correction as a signal imputation problem. While the methods developed in this chapter apply in principle to any pulsative signal, the focus is exclusively on the ECG given its direct relevance to the cardiac electrophysiology applications of this thesis.

To promote the development of imputation methods for pulsative signals, M. Xu et al. ([2022](#)) formulated a benchmark task in which ECG signals are subject to two categories of missing data: transient (short-interval) gaps due to information packet loss, and extended (long-interval) gaps due to connectivity issues. Their comprehensive study establishes a state-of-the-art deterministic solution for signal imputation using a transformer-based architecture. Two characteristics of pulsative signals are central to the imputation problem. First, these signals exhibit quasi-periodic structure with specific morphologies (e.g. the QRS complex in the ECG), which implies a strong template pattern across beats. Second, these morphologies vary over time and across populations:

different waveform characteristics are exhibited by healthy subjects and patients with different pathologies. These variations are driven by diverse factors including age, health condition, and recording time, yet this contextual information is typically unavailable during imputation.

DDPMs [J. Ho, Jain, and Abbeel, 2020; Y. Song et al., 2021; D. Kingma et al., 2021] have emerged as powerful probabilistic generative models, demonstrating strong performance in generation quality [Z. Chen, Y. Wu, et al., 2022; Whang et al., 2022] and mode coverage [Y. Song et al., 2021; Xiao, Kreis, and Vahdat, 2022]. A diffusion model comprises a forward process that progressively adds Gaussian noise to data samples until they reach a standard normal distribution, and a reverse process that iteratively generates data from this prior by learning the score function of the noisy distribution. Diffusion models have demonstrated superior results in time series imputation and forecasting [Tashiro et al., 2021; Alcaraz and Strodthoff, 2023], yet their potential for physiological signal reconstruction, where domain-specific structure can provide informative conditioning, has not been fully exploited.

This chapter introduces PULSEDIFF, a diffusion model conditioned on an augmented waveform template prior designed to exploit the quasi-periodic structure of the ECG to improve imputation accuracy. PULSEDIFF addresses three challenges through corresponding innovations (see Fig. 4.1):

1. *Template extraction*: A subject-level waveform template is extracted from observed values using matched filtering and beat detection, providing a personalised prior that captures the individual’s characteristic waveform morphology.
2. *Stochastic augmentation*: Beat-level random shifts in position and amplitude are applied to the template prior, modelling the uncertainty in beat detection and the natural variability in pulse timing and morphology across cardiac cycles.
3. *Rhythm confidence scoring*: A confidence score quantifies the regularity of the observed rhythm, determining whether the template prior should be extended to impute missing beat locations or restricted to detected beats only, ensuring that the prior is provided cautiously for subjects with irregular rhythms.

PULSEDIFF is agnostic to the choice of baseline diffusion architecture and can be applied on top of any conditional diffusion model. Validated on the PTB-XL dataset [Wagner et al., 2020], PULSEDIFF improves the reconstruction accuracy of two baseline diffusion models, CSDI [Tashiro et al., 2021] and SSSD^{S4} [Alcaraz and Strodthoff, 2023], with PULSEDIFF-SSSD^{S4} outperforming the leading deterministic method for transient missing data and remaining competitive for

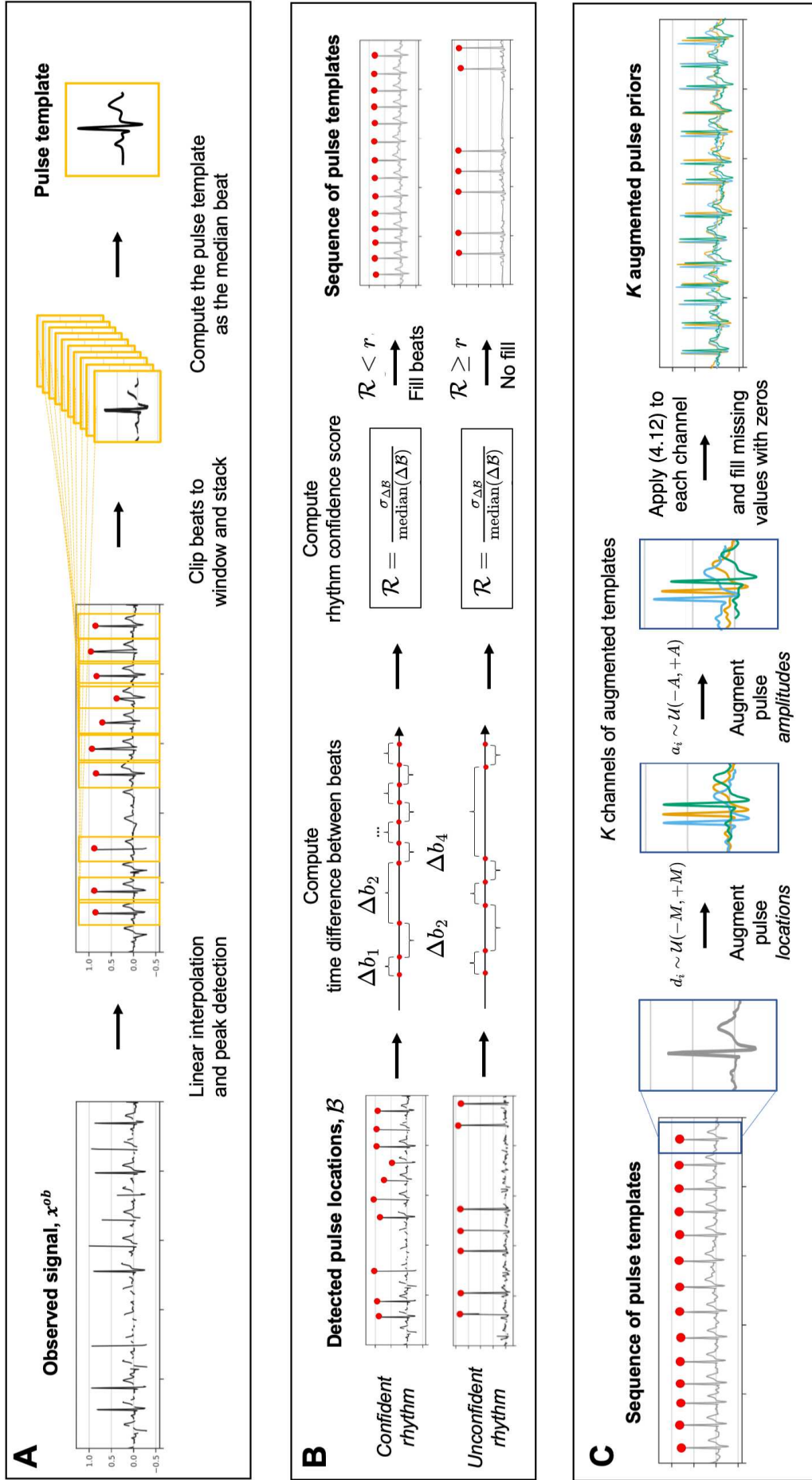


Figure 4.1: The proposed PULSEDIFF methodology to improve diffusion models for pulsative signal imputation. A) Calculation of the pulse template for each subject. B) Formation of the pulse prior using the rhythm confidence score to / to not fill missing beats. C) Augmentation of the pulse prior by applying random shifts in pulse location and amplitude.

extended gaps. Fig. 4.2 illustrates representative imputation results.

This chapter serves a dual purpose within the thesis. First, it demonstrates that domain-specific temporal priors can substantially improve imputation quality when the target signal exhibits quasi-periodic structure, providing a proof-of-concept for temporal prior-guided generative modelling of physiological signals. Second, it highlights the limitations of purely temporal approaches: the template extraction, rhythm confidence scoring, and beat-level augmentation developed here all rely fundamentally on the repeating morphology of the cardiac cycle. This chapter operates at the body-surface scale (Sec. 3.1), where the ECG captures the aggregate global electrical signature of the heart as a univariate time series. During atrial fibrillation, however, the electrical activity at the organ scale is characterised by chaotic, spatiotemporally complex dynamics (Sec. 3.3) that violate these quasi-periodicity assumptions. The absence of a repeating activation template, combined with significant beat-to-beat variability and multiple competing wavefronts, means that temporal priors alone cannot support the reconstruction of global fibrillation dynamics from sparse measurements. This limitation motivates the incorporation of spatial structure as a relational inductive bias in Chapter 5, which moves to the organ scale and exploits graph-based representations of cardiac tissue connectivity to enable reconstruction even when temporal patterns are irregular.

4.2 Preliminaries

4.2.1 ECG Imputation

The benchmark task for ECG imputation in the mHealth setting was established by M. Xu et al. (2022), who frame the problem within the broader class of pulsative signal imputation and provide a comprehensive study spanning classical signal processing methods (mean filling [Le, Beuran, and Tan, 2018], linear interpolation [Xinzheng Dong et al., 2019], and Fourier-based imputation [Rahman et al., 2015]) and deep learning approaches including RNN-based imputation via BRITS [Cao et al., 2018], transformer architectures [Bansal, Deshpande, and Sarawagi, 2021; Vaswani et al., 2017; S. Li et al., 2019], and generative adversarial network-based methods [J. Ho and Ermon, 2016; Yukai Liu et al., 2019]. These general-purpose imputation methods do not exploit the quasi-periodic structure or characteristic morphologies of the ECG, and their reconstruction performance is accordingly limited.

To address this, M. Xu et al. (2022) designed a bottleneck dilated convolutional transformer (BDC Trans) that achieves state-of-the-art deterministic performance by incorporating architectural inductive biases suited to the temporal structure of pulsative signals. However, probabilistic generative models are not considered in their study, leaving open the question of whether the stochastic

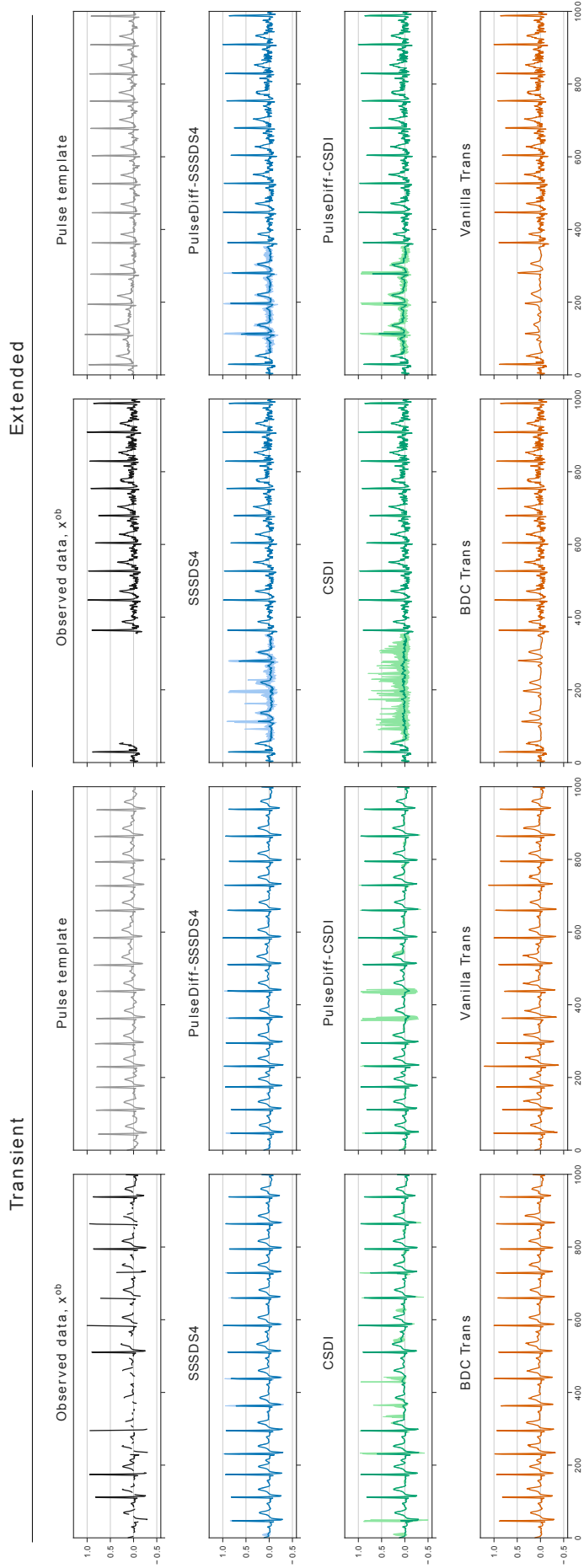


Figure 4.2: Comparison of imputation results on ECG waveforms from the PTB-XL dataset with 30% transient and extended missing values. PULSEDIFF improves performance of conditioning denoising diffusion probabilistic models CSDI [Tashiro et al., 2021] and SSSD^{S4} [Alcaraz and Strodthoff, 2023] by leveraging a pulse template as additional conditioning information. Imputation performance with PULSEDIFF-SSSD^{S4} also beats the state-of-the-art transformer model, BDC Trans, for transient missing values and is comparable for extended missing values.

variability inherent in ECG morphology across beats and subjects can be better captured by models that learn the conditional distribution over missing values rather than a single point estimate. This chapter addresses this gap by establishing probabilistic baselines for the ECG imputation benchmark and introducing domain-specific conditioning techniques that improve diffusion-based reconstruction.

4.2.2 Diffusion Models for Time Series Imputation

The theoretical background of this thesis introduces DDPMs as probabilistic extensions of neural sequence models that learn to sample from the full conditional distribution over missing or future values (Sec. 2.1.5). This section reviews their application to time series imputation in greater detail, as diffusion models form the methodological basis for PULSEDIFF. The mathematical formulation of the conditional diffusion process used in this chapter is presented in Sec. 4.3.1.

DDPMs [J. Ho, Jain, and Abbeel, 2020; Y. Song et al., 2021] define a forward Markov chain that progressively corrupts data with Gaussian noise until the signal distribution converges to a standard normal, and a learned reverse process that iteratively denoises samples to recover the data distribution. Training proceeds by optimising a variational bound on the data log-likelihood, which reduces in practice to a denoising score-matching objective [J. Ho, Jain, and Abbeel, 2020; Y. Song et al., 2021]. For imputation and forecasting, conditional formulations generate missing values by conditioning the reverse process on observed data. CSDI [Tashiro et al., 2021] implements this by training a score-based diffusion model with a self-supervised masking strategy, conditioning on observed values through a transformer-based architecture and demonstrating strong performance across traffic, air quality, and healthcare datasets. SSSD^{S4} [Alcaraz and Strodthoff, 2023] replaces the attention-based backbone with structured state-space models [Goel et al., 2022] to improve the modelling of long-range temporal dependencies while reducing computational cost. Both methods treat time series imputation as a general-purpose conditional generation problem and do not exploit domain-specific signal structure in their conditioning mechanism.

Applications of diffusion models to ECG and bioelectrical signal processing remain limited. DeScoD-ECG [H. Li et al., 2022] applies score-based diffusion to ECG denoising, while Alcaraz and Strodthoff (2023) explore unconditional ECG generation. Neither work addresses the imputation problem or exploits the quasi-periodic structure and subject-specific morphological variations of the ECG waveform. PULSEDIFF fills this gap by conditioning the diffusion process on an augmented template prior that encodes both the temporal structure and subject-level morphology of the ECG. The approach is agnostic to the choice of baseline diffusion architecture, as demonstrated by its application to both CSDI [Tashiro et al., 2021] and SSSD^{S4} [Alcaraz and Strodthoff,

2023] in this chapter.

Since the presentation of this chapter’s work [Jenkins, Z. Chen, et al., 2023], additional diffusion-based methods for ECG processing have appeared. DiffECG [Neifar et al., 2024] proposes a versatile diffusion model for single-heartbeat ECG generation, imputation, and forecasting on the MIT-BIH dataset, conditioning the reverse process on spectrogram representations. However, their imputation operates at the scale of individual cardiac cycles (270 samples) rather than full multi-second recordings, and does not exploit subject-level morphological priors or the quasi-periodic temporal structure leveraged by PULSEDIFF.

4.3 Methodology

4.3.1 Problem Formulation

ECG imputation as conditional generation. A single-channel ECG recording is a discrete-time signal $\{x[n]\}_{n=0}^{L-1}$ sampled at rate f_s (Sec. 2.1), collected into the vector $\mathbf{x} = [x[0], \dots, x[L-1]]^\top \in \mathbb{R}^L$. In the mHealth setting, subsets of samples may be missing due to electrode artefacts or connectivity loss. An observation mask, $\mathbf{m} \in \{0, 1\}^L$, indicates which samples are observed ($m_l = 1$) and which are missing ($m_l = 0$). The mask defines the observed portion of the signal as $\mathbf{m} \odot \mathbf{x} \in \mathbb{R}^L$, where $\odot$ denotes the Hadamard (element-wise) product and missing samples are set to zero. The observed and missing subsets of samples are denoted as $\mathbf{x}^{\text{ob}}$ and $\mathbf{x}^{\text{ta}}$, respectively, where the superscript “target” denotes the portion of the waveform to be imputed. Together, these form the complete signal, $\mathbf{x} = \mathbf{x}^{\text{ob}} + \mathbf{x}^{\text{ta}}$, such that the goal of imputation is to recover $\mathbf{x}^{\text{ta}}$ from $\mathbf{x}^{\text{ob}}$ to reconstruct the full waveform.

Deterministic imputation methods learn a mapping, $\hat{\mathbf{x}}^{\text{ta}} = f_{\boldsymbol{\theta}}(\mathbf{x}^{\text{ob}})$, that produces a single point estimate of the missing values. Probabilistic imputation instead aims to approximate the conditional distribution, $p(\mathbf{x}^{\text{ta}} | \mathbf{x}^{\text{ob}})$, with a parametric model, $p_{\boldsymbol{\theta}}(\mathbf{x}^{\text{ta}} | \mathbf{x}^{\text{ob}})$, enabling both reconstruction and uncertainty quantification through repeated sampling from the learned distribution. In this chapter each recording is represented by the tuple $\{\mathbf{x}, \mathbf{m}, \mathbf{s}\}$, where $\mathbf{s} = \{n\}_{n=0}^{L-1} \in \mathbb{R}^L$ stores the timestamps associated with each sample, following the notation of Tashiro et al. (2021).

Denoising diffusion probabilistic models. DDPMs [J. Ho, Jain, and Abbeel, 2020] define a generative model through two coupled stochastic processes: a forward (noising) process that progressively corrupts data into Gaussian noise, and a reverse (denoising) process that learns to invert this corruption.

Forward process. Starting from the data $\mathbf{x}_0 \sim q(\mathbf{x}_0)$, the forward process defines a Markov chain that adds Gaussian noise at each diffusion step, $\tau = 1, \dots, T$, according to a predetermined variance schedule $0 < \beta_1 < \dots < \beta_T < 1$,

$$q(\mathbf{x}_\tau | \mathbf{x}_{\tau-1}) = \mathcal{N}(\mathbf{x}_\tau; \sqrt{1 - \beta_\tau} \mathbf{x}_{\tau-1}, \beta_\tau \mathbf{I}). \quad (4.1)$$

The index τ denotes the diffusion step, distinguishing it from the discrete-time signal index, n , used elsewhere in this thesis (e.g. [Sec. 2.1](#)). Defining $\alpha_\tau := 1 - \beta_\tau$ and $\bar{\alpha}_\tau := \prod_{s=1}^\tau \alpha_s$, the marginal at any diffusion step can be expressed in closed form as

$$q(\mathbf{x}_\tau | \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_\tau; \sqrt{\bar{\alpha}_\tau} \mathbf{x}_0, (1 - \bar{\alpha}_\tau) \mathbf{I}), \quad (4.2)$$

which allows direct sampling of any noised version $\mathbf{x}_\tau = \sqrt{\bar{\alpha}_\tau} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_\tau} \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. As $T \rightarrow \infty$, the forward process drives the data distribution toward the isotropic Gaussian $\mathcal{N}(\mathbf{0}, \mathbf{I})$.

Reverse process. The generative (reverse) process starts from $p(\mathbf{x}_T) = \mathcal{N}(\mathbf{0}, \mathbf{I})$ and iteratively denoises through learned Gaussian transitions,

$$p_{\boldsymbol{\theta}}(\mathbf{x}_{0:T}) = p(\mathbf{x}_T) \prod_{\tau=1}^T p_{\boldsymbol{\theta}}(\mathbf{x}_{\tau-1} | \mathbf{x}_\tau), \quad (4.3)$$

where the subscript notation $\mathbf{x}_{0:T}$ denotes the sequence $(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_T)$ across all diffusion steps, and each reverse step is parameterised as

$$p_{\boldsymbol{\theta}}(\mathbf{x}_{\tau-1} | \mathbf{x}_\tau) = \mathcal{N}(\mathbf{x}_{\tau-1}; \boldsymbol{\mu}_{\boldsymbol{\theta}}(\mathbf{x}_\tau, \tau), \sigma_\tau^2 \mathbf{I}). \quad (4.4)$$

Following the $\boldsymbol{\varepsilon}$ -prediction parameterisation of J. Ho, Jain, and Abbeel (2020), the mean is expressed as

$$\boldsymbol{\mu}_{\boldsymbol{\theta}}(\mathbf{x}_\tau, \tau) = \frac{1}{\sqrt{\alpha_\tau}} \left(\mathbf{x}_\tau - \frac{\beta_\tau}{\sqrt{1 - \bar{\alpha}_\tau}} \boldsymbol{\varepsilon}_{\boldsymbol{\theta}}(\mathbf{x}_\tau, \tau) \right), \quad (4.5)$$

where $\boldsymbol{\varepsilon}_{\boldsymbol{\theta}}(\mathbf{x}_\tau, \tau)$ is a neural network trained to predict the noise $\boldsymbol{\varepsilon}$ that was added in the forward process. The variance is set to the schedule-dependent constant $\sigma_\tau^2 = \frac{(1 - \bar{\alpha}_{\tau-1})}{(1 - \bar{\alpha}_\tau)} \beta_\tau$.

Training objective. The model is trained by maximising a variational lower bound on the data log-likelihood, which simplifies to a denoising objective [J. Ho, Jain, and Abbeel, 2020],

$$\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}_{\mathbf{x}_0, \boldsymbol{\varepsilon}, \tau} \left[\left\| \boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}_{\boldsymbol{\theta}} \left(\sqrt{\bar{\alpha}_\tau} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_\tau} \boldsymbol{\varepsilon}, \tau \right) \right\|_2^2 \right], \quad (4.6)$$

where τ is sampled uniformly from $\{1, \dots, T\}$ and $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.

Conditional diffusion for ECG imputation. For the imputation task, the reverse process is conditioned on the observed values, $\mathbf{x}^{\text{ob}}$, and the observation mask, $\mathbf{m}$, so that the model generates only the missing portion of the signal while knowing which samples are valid observations and which are missing. The conditional reverse process becomes

$$p_{\boldsymbol{\theta}}(\mathbf{x}_{0:T}^{\text{ta}} | \mathbf{x}^{\text{ob}}, \mathbf{m}) = p(\mathbf{x}_T^{\text{ta}}) \prod_{\tau=1}^T p_{\boldsymbol{\theta}}(\mathbf{x}_{\tau-1}^{\text{ta}} | \mathbf{x}_{\tau}^{\text{ta}}, \mathbf{x}^{\text{ob}}, \mathbf{m}), \quad (4.7)$$

with the noise predictor now receiving the observed data and mask as additional inputs, $\boldsymbol{\varepsilon}_{\boldsymbol{\theta}}(\mathbf{x}_{\tau}, \tau, \mathbf{x}^{\text{ob}}, \mathbf{m})$, and the conditional training objective,

$$\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}_{\mathbf{x}_0, \boldsymbol{\varepsilon}, \tau} \left[\left\| \boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}_{\boldsymbol{\theta}} \left(\sqrt{\bar{\alpha}_{\tau}} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_{\tau}} \boldsymbol{\varepsilon}, \tau, \mathbf{x}^{\text{ob}}, \mathbf{m} \right) \right\|_2^2 \right]. \quad (4.8)$$

This conditional formulation is the starting point for PULSEDIFF, which extends it by providing the noise predictor with an additional template prior derived from the quasi-periodic structure of the ECG, as described in the following section.

4.3.2 PULSEDIFF: Template-Guided Diffusion for ECG Imputation

PULSEDIFF extends the conditional diffusion framework of [Sec. 4.3.1](#) by providing the noise predictor with a data-driven template prior derived from the quasi-periodic structure of the ECG. Three design challenges motivate the approach. First, ECG morphology varies across subjects, so a population-average template may be a poor prior for an individual recording; subject-level template extraction is therefore required. Second, the health condition of the subject is typically unknown during imputation, yet it affects the regularity of the cardiac rhythm and hence the reliability of predicted beat locations; a mechanism to assess rhythmic confidence is needed to ensure accurate template guidance. Third, even a well-constructed template will not align perfectly with the ground-truth waveform due to natural heart rate variability and beat detection uncertainty; stochastic augmentation of the template is therefore introduced to prevent the prior from misleading the generative process. These three components are described below and illustrated in [Fig. 4.1](#).

Template extraction. For each individual recording, $\{\mathbf{x}, \mathbf{m}, \mathbf{s}\}$, a subject-level waveform template, $\mathbf{p} \in \mathbb{R}^{L_b}$ (where $L_b < L$ is the template length in samples, corresponding to one cardiac cycle), is extracted from the observed values, $\mathbf{x}^{\text{ob}}$. The extraction proceeds in two stages, using matched filtering to detect beat locations and median averaging to construct the template waveform.

Since signal processing algorithms for beat detection require a contiguous signal, the observed

values, $\mathbf{x}^{\text{ob}}$, are first linearly interpolated to produce $\mathbf{x}^{\text{ob-l}} \in \mathbb{R}^L$, a sequence with all missing values filled. An external template, $\mathbf{p}_{\text{ext}}$, is constructed from a small subset of $N_{\text{ext}} \ll N$ fully-observed recordings in the training set: a Christov detector [Christov, 2004] identifies QRS complex locations, each detection is shifted back by 250 ms to include the P-wave, and the median waveform across all detected beats is computed. This median-waveform external template serves as the matched filter for the first stage of beat detection. The matched filter is the optimal linear detector for a known waveform in additive noise, maximising the output signal-to-noise ratio [Oppenheim, Schafer, and Buck, 1999], which makes it well suited to the present setting where linear interpolation of missing samples introduces artefacts and suppresses characteristic waveform features such as the R-wave peak.

The Christov detector [Christov, 2004] was selected for QRS detection because it is well-validated, computationally lightweight, and widely adopted in the ECG processing literature. The template extraction pipeline is modular with respect to the choice of detector: any algorithm with comparable sensitivity and specificity for QRS complexes (e.g. the Pan–Tompkins algorithm [Pan and Tompkins, 1985]) could be substituted without requiring changes to the downstream conditioning framework.

Beat locations in $\mathbf{x}^{\text{ob-l}}$ are identified by cross-correlating the signal with the external template,

$$c[n] = \sum_m x^{\text{ob-l}}[n+m] p_{\text{ext}}[m], \quad (4.9)$$

and thresholding the resulting correlation at the 97th percentile, with a minimum separation of 400 ms between detections (corresponding to a maximum resting heart rate of 150 bpm). These thresholds were determined in consultation with a clinical expert. The detected beat locations are used to segment $\mathbf{x}^{\text{ob}}$, and the segments are overlaid to compute an intermediate subject-level template, $\mathbf{p}'$, as the median waveform.

To refine beat detection, the process is repeated using $\mathbf{p}'$ as the matched filter template, yielding a final ordered set of detected beat locations, $\mathcal{B} = (b_i)_{i=1}^{n_b}$, where b_i denotes the sample index of the i -th detected beat, n_b is the total number of detected beats, and $b_i < b_{i+1}$ for all i . The observed signal, $\mathbf{x}^{\text{ob}}$, is segmented at these refined locations and the median waveform computed to produce the final subject-level waveform template, $\mathbf{p}$.

In summary, the template extraction proceeds as follows:

1. The observed value sequence, $\mathbf{x}^{\text{ob}}$, is linearly interpolated to form $\mathbf{x}^{\text{ob-l}}$.
2. Beat locations, $\mathcal{B}$, are detected in $\mathbf{x}^{\text{ob-l}}$ via two-stage matched filtering using the Christov

detector [Christov, 2004].

3. The observed signal is segmented at the detected beat locations.
4. The segments are stacked and the median waveform is computed to yield the subject-level waveform template, $\mathbf{p}$.

Rhythm confidence score. A sequence-length template prior can be formed by placing subject-level waveform templates at predicted beat locations in regions where no beats were detected, provided the underlying rhythm is sufficiently regular. However, if the subject has an arrhythmia (e.g. atrial fibrillation) or significant heart rate variability, extending the template to undetected locations risks biasing the imputation. To determine when this extrapolation is appropriate, a rhythm confidence score is introduced that quantifies the regularity of the observed cardiac rhythm.

From the detected beat locations, $\mathcal{B}$, the ordered set of inter-beat intervals $\Delta\mathcal{B} = (\Delta b_1, \dots, \Delta b_{n_b-1})$ is computed, where $\Delta b_i = b_{i+1} - b_i$. The rhythm confidence score is defined as

$$\mathcal{R} = \frac{\sigma_{\Delta\mathcal{B}}}{\text{median}(\Delta\mathcal{B})}, \quad (4.10)$$

where $\sigma_{\Delta\mathcal{B}}$ is the standard deviation of the inter-beat intervals. This formulation resembles the coefficient of variation but uses the median rather than the mean to provide robustness to outliers caused by missing data. A low value of $\mathcal{R}$ indicates a regular rhythm, while a high value suggests irregular or arrhythmic activity.

The deliberate simplicity of this formulation is a design choice: the score is a lightweight, interpretable heuristic that avoids introducing additional learned parameters into the conditioning pipeline. Any sufficiently reliable measure of rhythmic regularity would serve the same gating role; the coefficient-of-variation formulation was chosen for its robustness and computational simplicity relative to the diffusion training loop.

When $\mathcal{R} < r$ (with the threshold fixed at $r = 1$ throughout this chapter), the detected beat sequence is extended by imputing additional beat locations at the median inter-beat interval, yielding an augmented ordered set of beat locations, $\mathcal{B}' = (b'_1, \dots, b'_{n'_b})$, with $n'_b > n_b$. The waveform template, $\mathbf{p}$, is then placed at all locations in $\mathcal{B}'$ to form the complete pulse prior. When $\mathcal{R} \geq r$, the template is placed only at the originally detected beat locations, $\mathcal{B}$. In both cases, inter-template gaps are

filled with zeros. Formally, the *fixed* pulse prior is

$$\mathbf{p}^{\text{fix}}[n] = \begin{cases} \sum_{i=1}^{n'_b} p[n - b'_i], & \text{if } \mathcal{R} < r, \\ \sum_{i=1}^{n_b} p[n - b_i], & \text{otherwise.} \end{cases} \quad (4.11)$$

Template augmentation. The fixed pulse prior assumes that detected (or imputed) beat locations and the template morphology exactly match the ground truth. In practice, four sources of uncertainty affect this alignment: uncertainty in beat detection, uncertainty in beat location imputation, variability in waveform amplitude, and natural heart rate variability. To prevent the prior from providing overconfident or misleading conditioning information, beat-level random perturbations in both temporal position and amplitude are applied.

An augmented pulse prior is constructed by applying independent random shifts to each beat,

$$\mathbf{p}^{\text{aug}}[n] = \begin{cases} \sum_{i=1}^{n'_b} (a_i + p[n - b'_i - d_i]), & \text{if } \mathcal{R} < r, \\ \sum_{i=1}^{n_b} (a_i + p[n - b_i - d_i]), & \text{otherwise,} \end{cases} \quad (4.12)$$

where $d_i \sim \mathcal{U}(-M, +M)$ is a random temporal shift applied to the i -th beat location and $a_i \sim \mathcal{U}(-A, +A)$ is a random amplitude offset applied to the i -th beat. Both are drawn independently for each beat. Inter-template gaps are again filled with zeros.

To improve coverage of the ground-truth waveform, K independent augmented priors $\mathbf{p}_1^{\text{aug}}, \dots, \mathbf{p}_K^{\text{aug}}$ are sampled, each computed from (4.12) with independent draws of $\{d_i\}$ and $\{a_i\}$. These are collected into the augmented template prior matrix, $\mathbf{P} \in \mathbb{R}^{K \times L}$, with the k -th row storing $\mathbf{p}_k^{\text{aug}}$.

Template-conditional diffusion. The augmented template prior, $\mathbf{P}$, is incorporated as additional conditioning information in the diffusion model (Fig. 4.3). Each recording is now represented by the tuple $\{\mathbf{x}, \mathbf{m}, \mathbf{s}, \mathbf{P}\}$, and the conditional reverse process of (4.7) is extended to

$$p_{\boldsymbol{\theta}}(\mathbf{x}_{0:T}^{\text{ta}} \mid \mathbf{x}^{\text{ob}}, \mathbf{m}, \mathbf{P}) = p(\mathbf{x}_T^{\text{ta}}) \prod_{\tau=1}^T p_{\boldsymbol{\theta}}(\mathbf{x}_{\tau-1}^{\text{ta}} \mid \mathbf{x}_{\tau}^{\text{ta}}, \mathbf{x}^{\text{ob}}, \mathbf{m}, \mathbf{P}), \quad (4.13)$$

where $\mathbf{x}_{0:T}^{\text{ta}} = (\mathbf{x}_0^{\text{ta}}, \mathbf{x}_1^{\text{ta}}, \dots, \mathbf{x}_T^{\text{ta}})$ denotes the sequence of target signal states across all diffusion steps and $\mathbf{x}_T^{\text{ta}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Each reverse step is parameterised as in (4.4), with the noise predictor now receiving both the observation mask and the template prior as additional inputs, $\boldsymbol{\varepsilon}_{\boldsymbol{\theta}}(\mathbf{x}_{\tau}, \tau, \mathbf{x}^{\text{ob}}, \mathbf{m}, \mathbf{P})$,

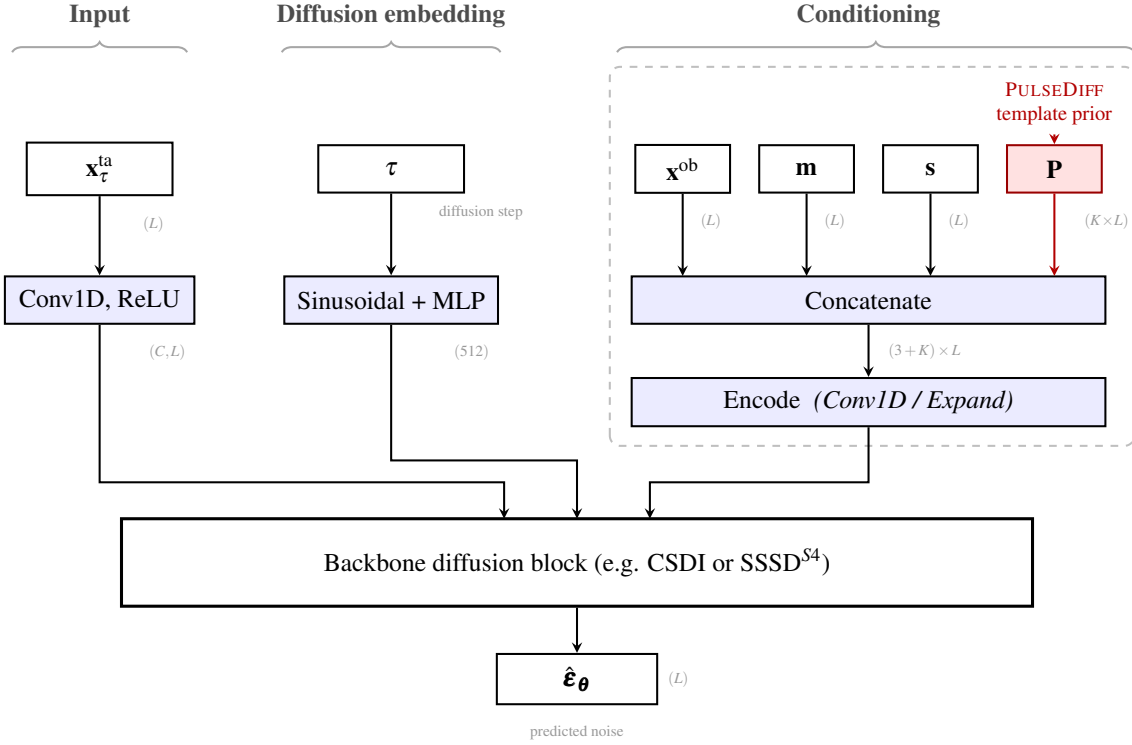


Figure 4.3: Overview of the PULSEDIFF conditioning mechanism. The noise predictor, $\boldsymbol{\varepsilon}_{\boldsymbol{\theta}}$, receives three input streams: the noised target signal, $\mathbf{x}_{\tau}^{\text{ta}}$, a sinusoidal diffusion step embedding, and conditioning information comprising the observed signal, $\mathbf{x}^{\text{ob}}$, the observation mask, $\mathbf{m}$, timestamps, $\mathbf{s}$, and the augmented template prior, $\mathbf{P} \in \mathbb{R}^{K \times L}$ (highlighted in red). The template prior is the contribution of PULSEDIFF; all other components are standard elements of the baseline CSDI [Tashiro et al., 2021] and SSSD^{S4} [Alcaraz and Strodthoff, 2023] architectures. The conditioning inputs are concatenated and encoded before being passed, alongside the encoded input and diffusion embedding, to the backbone diffusion block.

and the mean function given by

$$\boldsymbol{\mu}_{\boldsymbol{\theta}}(\mathbf{x}_{\tau}^{\text{ta}}, \tau \mid \mathbf{x}^{\text{ob}}, \mathbf{m}, \mathbf{P}) = \frac{1}{\sqrt{\alpha_{\tau}}} \left(\mathbf{x}_{\tau}^{\text{ta}} - \frac{\beta_{\tau}}{\sqrt{1 - \alpha_{\tau}}} \boldsymbol{\varepsilon}_{\boldsymbol{\theta}}(\mathbf{x}_{\tau}, \tau, \mathbf{x}^{\text{ob}}, \mathbf{m}, \mathbf{P}) \right). \quad (4.14)$$

The corresponding training objective is

$$\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}_{\mathbf{x}_0, \boldsymbol{\varepsilon}, \tau} \left[\left\| \boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}_{\boldsymbol{\theta}} \left(\sqrt{\alpha_{\tau}} \mathbf{x}_0 + \sqrt{1 - \alpha_{\tau}} \boldsymbol{\varepsilon}, \tau, \mathbf{x}^{\text{ob}}, \mathbf{m}, \mathbf{P} \right) \right\|_2^2 \right]. \quad (4.15)$$

Given the trained noise predictor, $\boldsymbol{\varepsilon}_{\boldsymbol{\theta}}$, imputation is performed by sampling $\hat{\mathbf{x}}_0^{\text{ta}}$ from the reverse process in (4.13), conditioned on the observed data, $\mathbf{x}^{\text{ob}}$, the observation mask, $\mathbf{m}$, and the augmented template prior matrix, $\mathbf{P}$. The complete reconstructed waveform is then $\hat{\mathbf{x}} = \mathbf{x}^{\text{ob}} + \hat{\mathbf{x}}_0^{\text{ta}}$.

4.4 Experiments

4.4.1 Experimental Setup

Dataset. All experiments were conducted on the PTB-XL dataset [Wagner et al., 2020], a large-scale open-source collection of clinical 12-lead ECG recordings sampled at 100 Hz with a duration of 10 s per recording. PTB-XL includes both healthy subjects and patients with co-occurring pathologies such as myocardial infarction and conduction disturbance, providing a diverse evaluation setting. Following the data split protocol of M. Xu et al. (2022), 8730 recordings were used for training, 2176 for validation, and 10,931 for testing. From the training set, 10 recordings were reserved for computing the external template, $\mathbf{p}_{ext}$ (Sec. 4.3.2), leaving 8720 for model training. Due to the high computational overhead of iterative sampling inherent to diffusion-based architectures, smaller fixed subsets were used for validation (64 samples) and testing (100 samples) to ensure feasible evaluation timelines while maintaining consistency across all baseline methods.

Missing value patterns. Following the PulseImpute benchmark [M. Xu et al., 2022], two categories of missing data were considered, corresponding to distinct failure modes in mHealth systems. Transient missing values simulate the sporadic loss of 50 ms information packets, while extended missing values model prolonged gaps due to sensor attachment issues. Both patterns are parameterised by a percentage controlling the proportion of removed samples. All models were trained with a fixed missing rate of 30%, the most common percentage observed in mHealth systems [M. Xu et al., 2022], and evaluated across missing rates from 10% to 50% at 10% increments to assess generalisation.

Evaluation protocol and metrics. Model performance was assessed using both reconstruction accuracy and downstream classification of the reconstructed waveforms. For the probabilistic diffusion models, each imputation is computed as the mean of 10 samples drawn from the learned conditional distribution, approximating $\mathbb{E}_{p_{\theta}}[\hat{\mathbf{x}}^{ta} | \mathbf{x}^{ob}]$, and reconstruction accuracy is quantified by the MSE between the imputed and ground-truth signals. All results report single evaluation values rather than mean $\pm$ standard deviation across random seeds, as the computational cost of diffusion model training and inference (each configuration requires 200 epochs on GPU with recursive application of the denoising process across diffusion steps) makes repeated training across multiple seeds prohibitive. The 10-sample averaging integrates over the stochasticity of the generative process, yielding stable point estimates for downstream evaluation. Deterministic baselines are evaluated using the pretrained models provided by M. Xu et al. (2022) under their original

protocol.

For downstream evaluation, imputed waveforms were passed through three pre-trained xResNet1d classifiers provided in M. Xu et al. (2022) to predict diagnostic labels (e.g. Wolff-Parkinson-White syndrome), form labels (e.g. inverted T-waves), and rhythm labels (e.g. atrial fibrillation), with performance measured by the macro area under the ROC curve (AUC). Additionally, a beat detection task was performed using the Stationary Wavelet Transform peak detector [Yun et al., 2022], with detected beat locations matched to ground-truth beats within a 50 ms tolerance window and evaluated using F1 score, precision, and recall, following the evaluation methodology of M. Xu et al. (2022). As a reference, the non-augmented pulse template prior (‘template’) was also evaluated by setting imputed values equal to the extracted template, demonstrating how PULSEDIFF leverages the prior beyond simple template filling.

Baselines. The deterministic baselines from the PulseImpute benchmark comprise BRITS with GAIL [J. Ho and Ermon, 2016], NAOMI [Yukai Liu et al., 2019], a vanilla transformer [Vaswani et al., 2017], Conv9 Trans [S. Li et al., 2019], and the state-of-the-art bottleneck dilated convolutional transformer (BDC Trans) [M. Xu et al., 2022]. For the transformer-based models, the open-source pretrained models provided by M. Xu et al. (2022) were used. As probabilistic baselines, CSDI [Tashiro et al., 2021] and SSSD^{S4} [Alcaraz and Strodthoff, 2023] were established by adapting their official implementations for single-channel ECG imputation. The template-guided variants are denoted PULSEDIFF-CSDI and PULSEDIFF-SSSD^{S4}, respectively.

Training details. For CSDI and PULSEDIFF-CSDI, the default architecture from Tashiro et al. (2021) was adopted: 4 residual layers, 64 residual channels, diffusion embedding dimension 128, and $T = 50$ diffusion steps with a quadratic noise schedule from $\beta_1 = 0.0001$ to $\beta_T = 0.5$. Since the task involves single-channel recordings, the feature transformer layer was removed and only the temporal transformer layer was retained, with embedding dimension 16 and 8 self-attention heads. Training proceeded for 200 epochs with batch size 16 using the Adam optimiser [D. P. Kingma and Ba, 2015], with a learning rate of 10^{-3} decayed to 10^{-4} at 75% and 10^{-5} at 90% of total epochs.

For SSSD^{S4} and PULSEDIFF-SSSD^{S4}, the architecture of Alcaraz and Strodthoff (2023) was followed: 36 residual layers, 256 residual and skip channels, diffusion embedding dimensions of 128, 256, and 256, and $T = 200$ diffusion steps with a linear noise schedule from $\beta_1 = 0.0001$ to $\beta_T = 0.02$. Training proceeded for 200 epochs with batch size 4 using the Adam optimiser with a fixed learning rate of 2×10^{-4} . All probabilistic models were trained on a single Nvidia Tesla

A30 GPU.

Hyperparameter selection. PULSEDIFF introduces three additional hyperparameters for tuning the augmented template prior: the magnitude of random shifts in beat location, M ; the magnitude of random shifts in beat amplitude, A ; and the number of augmented templates, K . Hyperparameter search was conducted via Bayesian optimisation over 50 iterations, each training for 25 epochs, with M and A selected continuously from $[0, 10]$ and K as an integer from $[1, 16]$. The optimal configuration was chosen to minimise MSE on the validation set. Bayesian optimisation identified the following configurations. For PULSEDIFF-CSDI: $M = 2$, $A = 0.01$, $K = 16$ (both transient and extended). For PULSEDIFF-SSSD^{S4}: $M = 1$, $A = 0.00$, $K = 8$ for transient missing values, and $M = 10$, $A = 0.00$, $K = 8$ for extended missing values.

4.4.2 Results

Reconstruction accuracy. Tables 4.1 and 4.2 report the MSE and downstream task performance for all methods at 30% transient and extended missing values, respectively. For transient missing data (Tab. 4.1), PULSEDIFF reduced the MSE of both baseline diffusion models, with PULSEDIFF-SSSD^{S4} achieving an MSE of 0.0021, the lowest among all probabilistic and deterministic methods. For extended missing data (Tab. 4.2), PULSEDIFF again improved both diffusion baselines, with PULSEDIFF-SSSD^{S4} achieving the best reconstruction accuracy among probabilistic models (MSE of 0.0192), comparable to the state-of-the-art deterministic result of BDC Trans [M. Xu et al., 2022]. Compared to the extracted pulse prior alone (Template rows in Tables 4.1 and 4.2), PULSEDIFF variants achieved substantially lower reconstruction error, confirming that the diffusion process leverages the template prior beyond simple template filling.

Fig. 4.4 extends this analysis across missing rates from 10% to 50%. PULSEDIFF variants consistently reduced the MSE of their respective baselines across all percentages tested, demonstrating that the template prior provides a robust conditioning signal that generalises across varying degrees of data loss.

Qualitative assessment. Fig. 4.2 presents representative imputation results for two subjects at 30% transient and extended missing values. By leveraging the template prior, PULSEDIFF variants improved the reconstruction quality of both CSDI and SSSD^{S4} by a considerable margin. Compared to CSDI, PULSEDIFF-CSDI generated substantially more realistic waveforms in regions of missing beats. Compared to SSSD^{S4}, PULSEDIFF-SSSD^{S4} recovered a greater extent of the waveform morphology and achieved improved estimation of beat amplitudes. The template prior

Table 4.1: Imputation performance on the PTB-XL dataset with 30% transient missing values. MSE quantifies reconstruction accuracy ($\downarrow$); classification AUC and beat detection metrics evaluate downstream task performance ($\uparrow$). Best overall results in **bold**. PULSEDIFF values are underlined where they improve on their baseline diffusion model, and colour-coded by relative improvement: **green** ($>1\%$), **amber** ($\pm 1\%$), **red** ($<-1\%$).

		CLASSIFICATION			BEAT DETECTION		
Model	MSE	Rhy.	Form	Diag.	F1	Prec.	Rec.
<i>Deterministic</i>							
BRITS+GAIL	.0346	.867	.781	.798	.464	.659	.358
Template	.0215	.937	.743	.842	.685	.747	.633
NAOMI	.0125	.939	.841	.853	.771	.761	.781
Van. Trans	.0081	.937	.821	.867	.838	.818	.859
Conv9 Trans	.0081	.924	.831	.873	.817	.804	.830
BDC Trans	.0043	.943	.813	.875	.900	.893	.906
<i>Probabilistic</i>							
CSDI	.0091	.936	.828	.861	.763	.839	.700
PULSEDIFF-CSDI	<u>.0065</u> _{+29%}	<u>.951</u> _{+1.6%}	<u>.831</u> _{+0.4%}	<u>.874</u> _{+1.5%}	<u>.851</u> _{+12%}	<u>.878</u> _{+4.6%}	<u>.825</u> _{+18%}
SSSD ^{S4}	.0023	.951	.845	.873	.930	.946	.916
PULSEDIFF-SSSD ^{S4}	<u>.0021</u> _{+8.7%}	<u>.955</u> _{+0.4%}	<u>.838</u> _{-0.8%}	<u>.882</u> _{+1.0%}	<u>.940</u> _{+1.1%}	<u>.963</u> _{+1.8%}	<u>.919</u> _{+0.3%}

also produced more informed and tightly bounded confidence intervals for the imputed values, reflecting the additional structural information provided to the diffusion process. Both PULSEDIFF variants produced more realistic pulse waveforms than the deterministic baselines BDC Trans [M. Xu et al., 2022] and Vanilla Trans [Vaswani et al., 2017].

Downstream task performance. Tables 4.1 and 4.2 also report downstream task performance. For transient missing values (Tab. 4.1), PULSEDIFF improved CSDI across all evaluation metrics and SSSD^{S4} on most metrics. PULSEDIFF-SSSD^{S4} achieved the best performance among all methods on five of six downstream metrics. For extended missing values (Tab. 4.2), PULSEDIFF improved CSDI by a substantial margin and achieved comparable performance to SSSD^{S4}. However, the advantage of diffusion models over deterministic baselines was less pronounced for extended gaps: BDC Trans [M. Xu et al., 2022] retained the best overall downstream performance in this setting. A likely explanation is that diffusion models are trained to estimate Gaussian noise rather than directly capturing temporal dependencies in the signal, and for extended gaps, the generated samples may deviate from the ground truth even under strong conditioning, whereas deterministic models optimised with a data-domain loss maintain closer fidelity.

Fig. 4.5 extends this analysis across missing rates from 10% to 50%. PULSEDIFF-CSDI demon-

Table 4.2: Imputation performance on the PTB-XL dataset with 30% extended missing values. MSE quantifies reconstruction accuracy ($\downarrow$); classification AUC and beat detection metrics evaluate downstream task performance ($\uparrow$). Best overall results in **bold**. PULSEDIFF values are underlined where they improve on their baseline diffusion model, and colour-coded by relative improvement: **green** ($>1\%$), **amber** ($\pm 1\%$), **red** ($<-1\%$).

		CLASSIFICATION			BEAT DETECTION		
Model	MSE	Rhy.	Form	Diag.	F1	Prec.	Rec.
<i>Deterministic</i>							
BRITS+GAIL	.0536	.887	.771	.801	.077	.076	.078
NAOMI	.0400	.910	.774	.840	.159	.165	.154
Template	.0363	.928	.837	.855	.470	.477	.464
Conv9 Trans	.0307	.911	.773	.814	.024	.900	.012
Van. Trans	.0248	.898	.774	.797	.458	.616	.365
BDC Trans	.0172	.960	.791	.807	.658	.716	.609
<i>Probabilistic</i>							
CSDI	.0306	.884	.767	.821	.033	.258	.018
PULSEDIFF-CSDI	<u>.0236</u> _{+23%}	<u>.927</u> _{+4.9%}	<u>.811</u> _{+5.7%}	<u>.840</u> _{+2.3%}	<u>.496</u> _{>100%}	<u>.603</u> _{>100%}	<u>.421</u> _{>100%}
SSSD ^{S4}	.0197	.933	.804	.864	.601	.722	.515
PULSEDIFF-SSSD ^{S4}	<u>.0192</u> _{+2.5%}	<u>.932</u> _{-0.1%}	<u>.803</u> _{-0.1%}	<u>.853</u> _{-1.3%}	<u>.608</u> _{+1.2%}	<u>.760</u> _{+5.3%}	<u>.507</u> _{-1.6%}

strated consistent improvements over CSDI across both transient and extended missing values at all percentages. PULSEDIFF-SSSD^{S4} improved upon SSSD^{S4} for transient missing values and performed comparably for extended missing values.

4.4.3 Ablation Study

To isolate the contribution of each proposed component, an ablation study was conducted by progressively adding components to the baseline CSDI until arriving at the full PULSEDIFF-CSDI model. The ablation was performed on the CSDI backbone as an insight study into the role of each conditioning component. CSDI was selected because its transformer-based architecture is representative of the predominant backbone design in diffusion models across domains [Vaswani et al., 2017; Tashiro et al., 2021], and because its lower computational cost relative to SSSD^{S4} permitted the multiple training runs required for systematic ablation within feasible timelines. For the same reason, all ablated models were trained for 50 epochs on data with 30% transient missing values, using the same hyperparameter configuration identified for PULSEDIFF-CSDI. To ensure balanced evaluation across the diverse pathologies in PTB-XL, the evaluation dataset was selected using stratified sampling with respect to the available class labels.

The ablation builds from the baseline CSDI by progressively introducing components: (i) add a

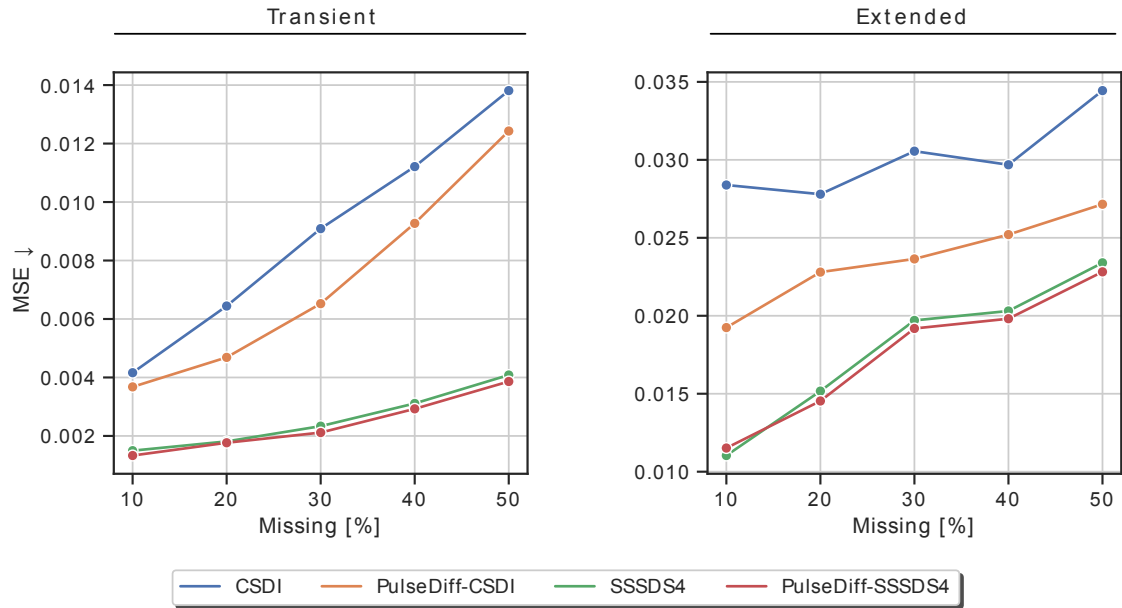


Figure 4.4: Reconstruction accuracy (MSE) of probabilistic diffusion models for different percentages of transient and extended missing values. PULSEDIFF variants consistently reduce the MSE of their respective baselines across all missing rates tested.

single fixed template prior, (ii) introduce the rhythm confidence score, (iii) increase the number of templates from 1 to $K = 16$ with beat-level stochastic augmentation, and (iv) combine augmentation with the rhythm confidence score to form the full PULSEDIFF-CSDI model.

Tab. 4.3 reports the results. The dominant effect is the introduction of any template prior: adding a single fixed prior to CSDI reduced MSE from .0177 to .0143 and improved all downstream metrics, confirming that providing the diffusion model with structural information about the expected waveform morphology yields substantial gains. Building on this, each proposed component individually improves upon the single fixed prior. Comparing CSDI with a single fixed prior against the same configuration with the rhythm confidence score, introducing the score reduced MSE from .0143 to .0122 and improved classification AUC across all three tasks, validating the utility of adaptive prior strength based on rhythmic regularity. Similarly, comparing the single fixed prior against 16 augmented priors (without the score), stochastic augmentation reduced MSE to .0120 and improved downstream performance, confirming that the augmentation mechanism captures meaningful variation across cardiac cycles.

However, the full PULSEDIFF-CSDI model (combining augmentation and scoring) achieved slightly higher MSE (.0135) than either component in isolation, indicating a non-additive interaction between the two mechanisms. A likely explanation is that the rhythm confidence score restricts the template prior to detected beat locations for subjects with irregular rhythms, reducing the spatial extent of the conditioning signal. While this gating is beneficial in isolation, preventing mislead-

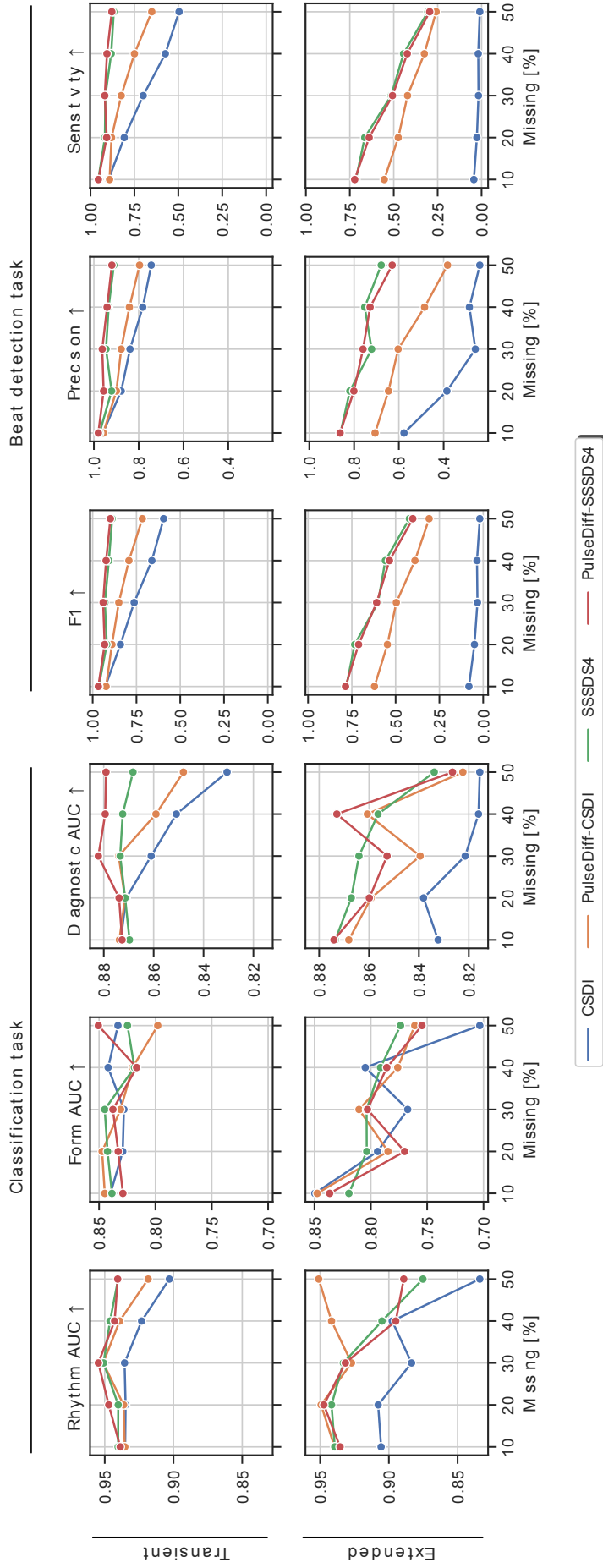


Figure 4.5: Performance of probabilistic diffusion models on classification and beat detection downstream tasks across different percentages of transient and extended missing values. PULSEDIFF-CSDI demonstrates consistent improvements over CSDI in both missing value types, while PULSEDIFF-SSSD^{S4} improves upon SSSD^{S4} for transient gaps and performs comparably for extended gaps.

Table 4.3: Ablation study for PULSEDIFF components evaluated on the PTB-XL dataset with 30% transient missing values. Each row progressively introduces components to the baseline CSDI model. Subscript percentages indicate relative improvement over the CSDI baseline, colour-coded as **green** ($>1\%$) and **amber** ($\pm 1\%$). Best result per column in **bold**.

Model	MSE	CLASSIFICATION			BEAT DETECTION		
		Rhy.	Form	Diag.	F1	Prec.	Rec.
CSDI	.0177	.886	.734	.712	.025	.030	.021
+ 1 fixed prior	.0143 _{+19%}	.925 _{+4.4%}	.741 _{+1.0%}	.721 _{+1.3%}	.053 _{>100%}	.046 _{+53%}	.061 _{>100%}
+ 1 prior + score	.0122 _{+31%}	.909 _{+2.6%}	.753 _{+2.6%}	.726 _{+2.0%}	.043 _{+72%}	.034 _{+13%}	.056 _{>100%}
+ 16 aug. prior	.0120 _{+32%}	.908 _{+2.5%}	.762 _{+3.8%}	.729 _{+2.4%}	.050 _{>100%}	.042 _{+40%}	.061 _{>100%}
PULSEDIFF-CSDI	.0135 _{+24%}	.908 _{+2.5%}	.752 _{+2.5%}	.725 _{+1.8%}	.037 _{+48%}	.032 _{+7%}	.045 _{>100%}

ing priors in arrhythmic recordings, it also reduces the number of locations at which the $K = 16$ augmented templates can provide informative conditioning. The augmentation benefit is therefore limited for precisely those subjects where imputation is most challenging. In effect, the score and augmentation mechanisms operate on competing objectives: the score acts conservatively by withholding prior information, while augmentation acts generously by providing diverse prior realisations.

Crucially, each component individually and substantially improves upon the baseline CSDI across all metrics, confirming that both the rhythm confidence score and the stochastic augmentation contribute meaningfully to imputation quality. The interaction effect suggests that replacing the hard-threshold gating of the rhythm confidence score ($\mathcal{R} > 1$) with a data-driven mechanism would better exploit the complementarity of the two components. For instance, classifier-free guidance [J. Ho and Salimans, 2022] could allow the model to learn when to rely on the full template prior versus the score-restricted version, effectively relaxing the fixed threshold to a soft, learned trade-off; a direction for future work.

4.5 Discussion

This chapter introduced PULSEDIFF, a template-guided conditioning framework for DDPMs applied to ECG imputation. The central idea is that the quasi-periodic structure of the ECG provides a powerful inductive bias for the imputation task: by extracting a subject-level waveform template from observed data and conditioning the reverse diffusion process on an augmented version of this template, the generative model is guided toward physiologically plausible reconstructions. The results on the PTB-XL dataset demonstrate that this domain-specific conditioning consistently im-

proves upon baseline diffusion models and achieves state-of-the-art reconstruction accuracy for transient missing data, while remaining competitive with the leading deterministic method for extended gaps.

Design for healthy and diseased conditions. A key design consideration of PULSEDIFF is its applicability across both healthy subjects and patients with cardiac pathologies. The PTB-XL dataset contains recordings spanning a broad range of conditions, from normal sinus rhythm to myocardial infarction, conduction disturbances, and rhythm disorders including atrial fibrillation. Rather than assuming a fixed waveform morphology, PULSEDIFF adapts to the individual through subject-level template extraction (Sec. 4.3.2), ensuring that the prior reflects the patient’s own cardiac characteristics. The stochastic augmentation mechanism (Eq. 4.3.2) further accounts for natural beat-to-beat variability in timing and amplitude, preventing the prior from imposing overconfident assumptions on recordings where such variability is pronounced.

The rhythm confidence score plays a particularly important role in this regard. For subjects with regular rhythms, the score permits the template to be extended to predicted beat locations in missing regions, providing informative conditioning across the full signal. For subjects with irregular rhythms, such as those with atrial fibrillation, the score restricts the template to detected beat locations only, expressing uncertainty in the temporal structure of the missing signal rather than risking misleading guidance. This adaptive gating reflects a fundamental principle: the value of a prior is contingent on its reliability, and it is preferable to withhold potentially erroneous information than to bias the reconstruction. The ablation study (Sec. 4.4.3) confirms that this cautious approach is effective, with the rhythm confidence score individually improving upon the fixed-prior baseline across all metrics. The non-additive interaction between the score and stochastic augmentation, where their combination yields slightly higher MSE than either component alone, highlights the tension between conservative and full conditioning strategies and suggests that a learned, soft gating mechanism such as classifier-free guidance [J. Ho and Salimans, 2022] could better exploit their complementarity.

Diffusion models for physiological signal imputation. Beyond the specific contributions of PULSEDIFF, this chapter establishes probabilistic baselines for the ECG imputation benchmark of M. Xu et al. (2022). Prior to this chapter, diffusion-based approaches had not been evaluated on the PulseImpute benchmark, and the question of whether probabilistic generative models could match or exceed the performance of deterministic architectures for physiological signal reconstruction remained open. The results show that SSSD^{S4} alone achieves reconstruction accuracy comparable to the state-of-the-art BDC Trans for transient missing data, and that PULSEDIFF-SSSD^{S4} sur-

passes it. This demonstrates that diffusion models are a viable and competitive class of methods for ECG imputation, with the additional benefit of providing confidence intervals through repeated sampling from the learned conditional distribution. The probabilistic formulation is particularly relevant for clinical applications, where a single point estimate may be insufficient and quantifying the uncertainty of the reconstruction can inform downstream decision-making.

The results also reveal that the advantage of diffusion models is less pronounced for extended missing data, where BDC Trans retains the best downstream classification performance. This asymmetry is consistent with the observation that diffusion models are trained to predict Gaussian noise rather than directly optimising a data-domain reconstruction loss: for short gaps surrounded by observed context, the conditioning signal constrains the reverse process effectively, but for extended gaps the accumulated stochasticity of the iterative denoising can produce samples that deviate from the ground truth despite strong conditioning. This suggests that hybrid approaches combining the distributional flexibility of diffusion models with the fidelity of data-domain objectives may be a fruitful direction for future work in this setting.

Limitations. Several limitations of this chapter should be acknowledged. The evaluation was conducted on a single dataset (PTB-XL) with synthetically generated missing value patterns. While the transient and extended missingness categories are designed to reflect realistic mHealth failure modes [M. Xu et al., 2022], the distribution of missing data in real wearable recordings may differ. The computational cost of diffusion models remains a practical concern: the iterative sampling procedure is substantially slower than a single forward pass through a deterministic model, and the 10-sample averaging used for evaluation compounds this cost. Advances in accelerated sampling for diffusion models [Tashiro et al., 2021; Lu et al., 2025] could mitigate this limitation. Finally, the template extraction pipeline depends on the quality of beat detection from partially observed signals. Although the two-stage matched filtering procedure is robust by design, severely corrupted recordings with very few observed samples may yield unreliable templates.

4.6 Summary and Relation to Thesis

This chapter demonstrated that patient-specific temporal priors can substantially improve the imputation of quasi-periodic physiological signals. PULSEDIFF achieves this by conditioning DDPMs on an augmented waveform template extracted from observed data, with a rhythm confidence score that adapts the prior strength to the regularity of the cardiac rhythm. The approach is agnostic to the choice of baseline diffusion architecture, as demonstrated by consistent improvements to both CSDI and SSSD^{S4}. These results address the problem of global-field ECG imputation from wear-

able devices at the patient level, recovering complete waveforms from recordings degraded by electrode artefacts and connectivity loss.

PULSEDIFF also serves a broader role within this thesis by establishing the promise and limitations of temporal priors for physiological signal reconstruction. The ECG at the body-surface scale (Sec. 3.1) aggregates the electrical activity of the entire heart into a univariate time series with a clear quasi-periodic structure: repeating P-wave, QRS-complex, and T-wave morphologies that recur with each cardiac cycle. This regularity is precisely what makes template-based conditioning effective. The rhythm confidence score quantifies when this regularity holds and, crucially, when it does not, providing an explicit mechanism to express uncertainty in the temporal prior for subjects whose rhythm is irregular. For the analysis of recordings that are exclusively atrial fibrillation, however, the rhythm will be inherently irregular and highly uncertain, and the temporal prior will be of limited value. The template-guided approach presented here fundamentally relies on the existence of a repeating waveform to extract and condition upon; when this assumption is violated, as in fibrillation, the temporal inductive bias loses relevance.

The transition from this chapter to Chapter 5 is motivated by the recognition that the reconstruction problem changes fundamentally when moving from the body-surface scale to the organ scale. At the organ scale, the electrical dynamics of the heart are no longer captured by a single univariate time series but by a collection of signals recorded at spatially distributed locations across the cardiac surface (Sec. 3.1). Although quasi-periodic structure exists at the cellular level in the form of the action potential, the signals observed at the tissue surface represent the superposition of electrical activity from thousands of cardiomyocytes. The cumulative effect of these superpositions, particularly during fibrillation when multiple wavefronts, re-entrant circuits, and phase singularities coexist (Sec. 3.3), produces dynamics that are difficult to characterise by temporal templates alone. The temporal prior that is so effective for the ECG breaks down at this scale.

This motivates two key methodological shifts in the subsequent chapter. First, the move from univariate to multivariate signals demands models that can process collections of spatially distributed time series simultaneously, representing a substantial increase in dimensionality. Second, the absence of reliable temporal priors necessitates alternative inductive biases that can constrain the reconstruction problem. Chapter 5 addresses both shifts by introducing relational inductive biases through graph-based representations of the atrial surface. By encoding the anatomical connectivity of cardiac tissue as a graph, spatial propagation constraints can guide reconstruction even when temporal patterns are irregular, enabling the recovery of global fibrillation dynamics from sparse catheter measurements. The spatiotemporal graph formulation exploits the fact that elec-

trical wavefronts propagate along tissue pathways, so that information at observed locations can be propagated to unobserved regions through learned message-passing operations on the graph. This spatial inductive bias replaces the temporal prior of PULSEDIFF as the primary structural constraint, addressing the limitations identified in this chapter. It is worth noting that diffusion models, demonstrated here for temporal imputation, may also offer advantages in the spatiotemporal reconstruction setting; the potential integration of generative modelling with graph-based spatiotemporal architectures is discussed at the close of [Chapter 5](#).

Chapter 5

Learning to Reconstruct Global Atrial Fibrillation Dynamics From Sparse Measurements

This chapter addresses the central reconstruction problem of this thesis: recovering global cardiac fibrillation dynamics from the sparse, localised measurements collected during routine catheter mapping. FIBMAP, the graph recurrent neural network proposed in this chapter, models the atrial surface as a graph and propagates information from observed to unobserved regions through learned message-passing operations. A training strategy that simulates clinical sequential mapping prepares the model for deployment on sparse data, while patient-specific and node-specific embedding parameters personalise reconstruction to individual electrophysiological characteristics. Validated on high-resolution optical mapping of rat ventricular fibrillation and clinical non-contact mapping of human atrial fibrillation, FIBMAP reconstructs global dynamics from 10% instantaneous surface coverage, substantially outperforming baselines in both signal fidelity and phase singularity detection. A supplemental translational showcase demonstrates reconstruction of whole-atrial dynamics from routine contact mapping recordings. Beyond its standalone contributions, this chapter provides a proof-of-concept for imputation mapping, while the model-dependent nature of its uncertainty estimates motivates the distribution-free conformal prediction framework developed in the subsequent chapter.

5.1 Introduction	93
5.2 Preliminaries	95
5.3 Methodology	95
5.4 Experiments	102
5.5 Results	113
5.6 Discussion	124
5.7 Summary and Relation to Thesis	128

This chapter is based on Jenkins, Cini, Barker, Sharp, et al. (2025).

5.1 Introduction

The preceding chapter developed PULSEDIFF, a template-guided diffusion model for ECG imputation whose design relies fundamentally on the quasi-periodic structure of the cardiac cycle (Chapter 4). This chapter moves to a fundamentally different setting in two respects. First, the focus shifts from the body-surface scale to the organ scale (Sec. 3.1), where the electrical dynamics of the heart are captured not by a single time series but by a collection of spatially distributed signals recorded across the cardiac surface. Second, the target application is AF, whose chaotic, spatiotemporally disorganised dynamics (Sec. 3.3) violate the quasi-periodicity assumption on which template-based conditioning relies, rendering the temporal prior of the previous form less relevant. In place of temporal priors, this chapter investigates relational inductive biases through graph-based spatiotemporal machine learning, a framework that naturally describes signals evolving over irregular spatial domains by encoding the anatomical connectivity of cardiac tissue as a graph and propagating information between neighbouring regions through learned message-passing operations.

As established in Chapter 3, AF is the most common cardiac arrhythmia and a major driver of global morbidity and mortality. Catheter ablation aims to restore sinus rhythm through targeted tissue destruction (Sec. 3.4), yet success rates remain as low as 50% at 5 years in persistent AF [Scherr et al., 2015], with alternative strategies such as linear lesions and targeted driver ablation failing to significantly improve outcomes [J. M. Lee et al., 2019; Verma et al., 2015; Vogler et al., 2015; Narayan, Krummen, et al., 2012; Wong et al., 2015]. A fundamental reason is that current approaches do not tailor interventions to the unique electrophenotypes within the spectrum of AF observed in patients (Sec. 3.3), leaving a one-size-fits-all treatment plan for a highly heterogeneous disorder [F. S. Ng et al., 2020].

The major limitation in implementing personalised ablation strategies is the inability to map AF effectively, specifically, the challenge of reconstructing global, whole-atrial dynamics on the endocardial surface [Deb et al., 2021; Cheniti et al., 2018; Boyle et al., 2019; Gillis et al., 2013; Roney et al., 2017]. As discussed in Sec. 3.5, sequential contact mapping provides high local signal fidelity but covers only a fraction of the atrial surface at any instant. While sequential recordings can be computationally stitched into coherent maps for stable organised arrhythmias [Özgül et al., 2023], the significant beat-to-beat variations in wavefront propagation during AF mean that non-contemporaneous, sparse recordings cannot be sensibly assembled to represent a coherent global map of the dynamics. Non-contact mapping systems achieve simultaneous whole-atrial coverage but rely on ill-posed inverse reconstruction, limiting spatial resolution and clinical adoption

(Fig. 3.5.1) [Narayan, Krummen, et al., 2012; Buch et al., 2016; Rudy and Burnes, 1999; Roney et al., 2017]. No existing mapping modality simultaneously delivers the spatial resolution, temporal simultaneity, and full-chamber coverage required to characterise the complex spatiotemporal dynamics of AF in clinical practice (Sec. 3.6).

Current machine learning approaches to improve AF mapping focus on signal annotation and classification from local electrogram recordings. For instance, deep learning can identify focal sources, triggers, and areas of spatiotemporal dispersion from electrograms to guide ablation [Liao et al., 2021; Seitz et al., 2022; Manongi, J. Kim, and Goldbarg, 2023]. These methods, while valuable for local annotation, cannot provide a complete global view of the wavefront dynamics that occur across the whole atria during AF.

To address this challenge, this chapter introduces *imputation mapping*, a graph machine learning [Bronstein, Bruna, LeCun, et al., 2017; Stanković, D. Mandic, et al., 2020c; Bacciu et al., 2020] approach designed to reconstruct complete global atrial dynamics from the sparse, localised sequential contact measurements collected during routine procedures. The proposed model, FIBMAP, utilises a GRNN architecture (Sec. 2.3.2) to model the sparse time-varying signals on the atrial surface, by representing the atrium as a graph where tissue regions are nodes connected by spatial adjacencies. By integrating the spatial connectivity of graphs with recurrent temporal modelling, FIBMAP learns the underlying spatiotemporal dynamics of cardiac fibrillation, enabling the propagation of information from observed to unobserved regions [Seo et al., 2018; Cini, Marisca, and Alippi, 2022]. Specifically, FIBMAP is designed to: (i) capture patient-specific fibrillation dynamics through learnable embedding parameters (Sec. 2.3.3); (ii) predict whole-atrial maps from routine contact measurements via spatiotemporal imputation; and (iii) efficiently generalise to new patients through rapid fine-tuning of only the embedding parameters.

To this end, FIBMAP provides a proof-of-concept for AF mapping that reconstructs global dynamics within the constraints of existing clinical equipment. By transforming sparse sequential measurements into comprehensive spatiotemporal visualisations, this approach demonstrates the capacity to reveal aspects of AF behaviour often hidden during sparse localised sampling. This facilitates the characterisation of patient-specific mechanisms, such as rotational activity or local drivers, directly from routine mapping data. This chapter addresses the longstanding challenge of visualising complex AF dynamics and establishes a foundation for more precise interventions tailored to individual arrhythmic patterns, as the field advances toward personalised AF management.

5.2 Preliminaries

FIBMAP builds on the STGNN framework developed in [Sec. 2.3.2](#), adopting a T&S GRNN architecture with the hybrid global-local embedding strategy motivated in [Sec. 2.3.3](#). This section situates the approach within the broader landscape of imputation methods for time series and spatiotemporal data.

Imputation methods span a wide spectrum, from classical matrix factorisation (MF) approaches [Salakhutdinov and Mnih, 2008; D. Lee and Seung, 2000] and their extensions with graph and temporal regularisation [Cai et al., 2010; H.-F. Yu, N. Rao, and Dhillon, 2016], to deep learning techniques employing RNNs [Cao et al., 2018], generative adversarial networks [Yoon, Jarrett, and Van der Schaar, 2019; Yukai Liu et al., 2019], transformers [Du, Côté, and Yan Liu, 2023], and denoising diffusion probabilistic models [Tashiro et al., 2021; Alcaraz and Strodthoff, 2023; Jenkins, Z. Chen, et al., 2023]. Among these, graph-based imputation methods that exploit relational structure for spatiotemporal missing data have achieved state-of-the-art results across a range of domains [Marisca, Cini, and Alippi, 2022; Cini, Marisca, and Alippi, 2022]. FIBMAP extends this line of work by introducing patient-specific and node-specific embeddings to address the heterogeneous dynamics of cardiac fibrillation, drawing on hybrid global-local STGNNs that augment shared parameters with learnable node embeddings [Cini, Marisca, Zambon, et al., 2024; X. Yin et al., 2022]. However, none of these methods have been applied to the reconstruction of cardiac fibrillation dynamics from sparse clinical measurements, where the combination of chaotic spatiotemporal activity, patient-specific substrates, and extreme sparsity presents distinct challenges.

5.3 Methodology

This study evaluates the FIBMAP imputation mapping framework through a multi-stage approach that progresses from development and validation to a supplemental translational showcase on clinical sequential contact mapping. The following sections describe the model architecture ([Sec. 5.3.2](#)) and the datasets used to evaluate the model ([Sec. 5.4.1](#)). A core proof-of-concept is established using two primary benchmarking datasets: foundational evaluation in ex vivo rat VF ($n = 17$) and a clinical cohort of human persistent AF ($n = 51$). As an extension beyond this core validation, a supplemental translational showcase is presented involving real samples of sequential contact mapping collected in patients. Finally, the training, fine-tuning, and validation protocols used to assess spatiotemporal reconstruction accuracy are detailed in [Sec. 5.4.2](#).

5.3.1 Problem Formulation

The chamber surface is modelled as a graph, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, with adjacency matrix, $\mathbf{A} \in \mathbb{R}^{N \times N}$, constructed by discretising patient-specific geometry into N nodes connected through surface triangulation, with adjacency remaining constant over time (see Fig. 5.1A). This graph representation captures the anatomical structure of the chamber surface, naturally defining the irregular domain through which electrical signals propagate, and provides a strong relational inductive bias that ensures information propagation is modelled according to physical tissue connectivity. Let $x_i[t] \in \mathbb{R}$ denote the scalar electrical signal at node i and time t , with $\mathbf{X}[t] \in \mathbb{R}^{N \times 1}$ representing the stacked signals across all N nodes and $\mathbf{X}[t:t+T]$ indicating the sequence over the time interval $[t, t+T)$. Due to sequential contact mapping, valid observations are sparse and their locations are indicated by a binary mask, $\mathbf{M}[t] \in \{0, 1\}^{N \times 1}$, where unity denotes valid measurements.

The objective of imputation mapping is to reconstruct the complete spatiotemporal signals, $\mathbf{X}[t:t+T]$, from the sparse measurements, estimating missing elements where $\mathbf{M}[t:t+T]$ is zero using the available contact measurements. Since clinical procedures measure only electrical signals while tissue properties influencing dynamics remain unmeasured, two sets of trainable embedding parameters are introduced to capture these unknown factors: a node embedding matrix, $\mathbf{V} \in \mathbb{R}^{N \times d_q}$, whose rows learn spatially varying tissue properties (e.g. conductivity) that influence local dynamics, and a patient embedding vector, $\mathbf{g} \in \mathbb{R}^{d_r}$, which captures global patient characteristics affecting AF behaviour (e.g. covariates), where d_q and d_r denote the node and patient embedding dimensions, respectively. These parameters provide node- and patient-level specialisation to the imputation model, following the hybrid global-local strategy motivated in Sec. 2.3.3, and alleviating the need to train separate models for each atrial region or patient. The complete dynamical graph representation of the data for a given patient is

$$\mathcal{D}[t:t+T] = \langle \mathbf{X}[t:t+T], \mathbf{M}[t:t+T], \mathbf{A}, \mathbf{V}, \mathbf{g} \rangle, \quad (5.1)$$

as illustrated in Fig. 5.1A. This graph-structured time series representation enables the use of STGNNs (Sec. 2.3.2) to efficiently learn global reconstruction from sparse observations.

5.3.2 FibMap: Architecture

FIBMAP is instantiated as a Bidirectional (Bi)-GRNN, a non-linear state-space model designed to propagate information from valid observations across space and time. The architecture follows the encode–process–decode framework presented in Sec. 2.3.2: two spatiotemporal encoder blocks process the sequence in both a forward (*fwd*) and backward (*bwd*) direction, and a de-

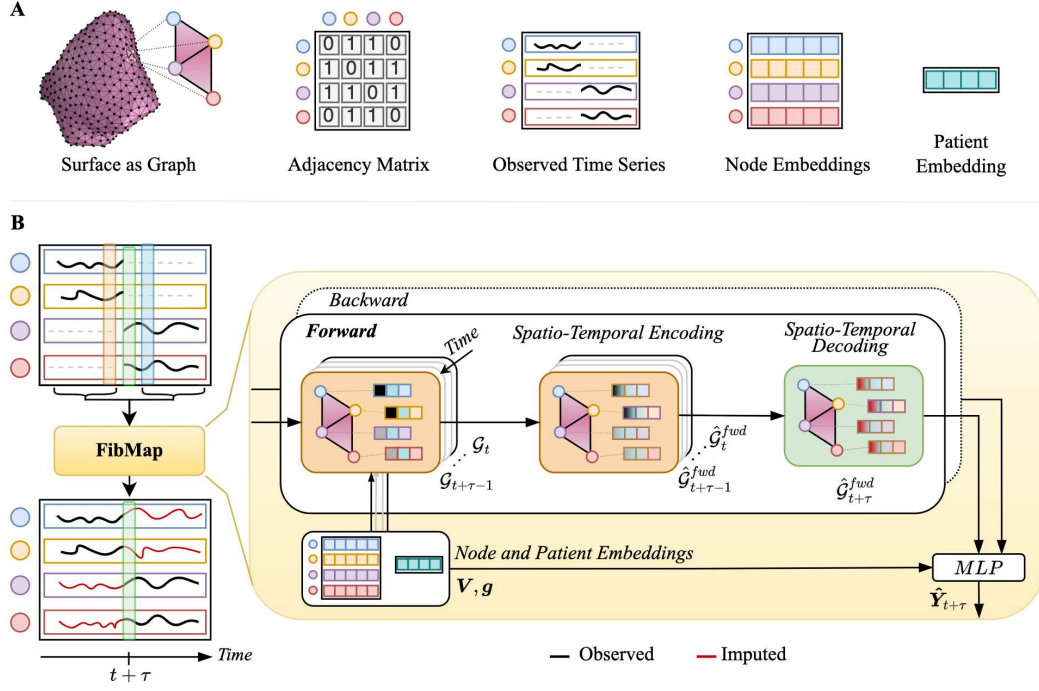


Figure 5.1: FIBMAP inputs and architecture. A) Model inputs: chamber surface discretised as a graph with adjacency matrix encoding node connectivity, observed time series at each node (sparse due to sequential contact mapping), learnable node embeddings capturing spatially varying tissue properties, and patient embedding capturing global characteristics. B) The bidirectional graph recurrent neural network architecture of FIBMAP. Input signals and embeddings are encoded via spatiotemporal processing, then decoded through a MLP to produce imputed signals (red) alongside observed measurements (black). The bidirectional structure propagates information both forwards and backwards through time, with hidden states exchanged between directions. Node and patient embeddings personalise both encoding and decoding operations.

coder combines representations from both directions to produce the final imputation. FIBMAP’s architecture extends the framework proposed in [Cini, Marisca, and Alippi, 2022] by introducing local (node) and global (patient) embedding parameters to address the AF mapping problem. These modifications enable generalisation across patients and efficient transfer to new patients via fine-tuning, following the hybrid global-local strategy discussed in Sec. 2.3.3.

To balance the trade-off between personalised reconstruction and generalisation across a spectrum of AF dynamics with a single model, FIBMAP is designed to perform personalised reconstruction by providing patient-specific parameters, $\mathbf{V}$ and $\mathbf{g}$, as additional input to the encoder and decoder components of the architecture. All other model parameters are shared across patients to learn the common aspects of the dynamics (such as the physics of the problem). The following begins by defining the spatiotemporal encoder block of the architecture, before explaining the decoder. Throughout, the notation conventions established in Sec. 2.3.2 are adopted: layer indices are denoted by superscript (l) , neighbourhood sets by $\mathcal{N}_i$, message and update functions by $\rho^{(l)}$ and $\gamma^{(l)}$,

and message-passing operators by $\text{MP}^{(l)}$ acting on the graph $\mathcal{G}$. Time indices are written in square brackets, $[t]$.

The Spatiotemporal Encoder

First, the observed data at time t is encoded as

$$\mathbf{H}^{(0)}[t] = \text{MLP}_{\text{enc}}([\mathbf{X}[t] \odot \mathbf{M}[t] \parallel \mathbf{M}[t] \parallel \mathbf{G} \parallel \mathbf{V}]), \quad (5.2)$$

where the symbols $\odot$ and $\parallel$ denote the Hadamard product and concatenation operator, respectively, and $\mathbf{G}^{(p)} \in \mathbb{R}^{N \times d_r}$ is a matrix where copies of $\mathbf{g}$ are stacked across nodes. The encoding step constructs a representation of the observed and missing values alongside the patient-specific parameters in a common embedding space, $\mathbf{H}^{(0)}[t] \in \mathbb{R}^{N \times d}$. Next, the embedded data is processed sequentially using a gated-GRNN, where the l -th layer is given by

$$\mathbf{Z}^{(l)}[t] = \mathbf{H}^{(l-1)}[t], \quad (5.3a)$$

$$\mathbf{R}^{(l)}[t] = \sigma(\text{MP}_r^{(l)}([\mathbf{Z}^{(l)}[t] \parallel \mathbf{H}^{(l)}[t-1]], \mathcal{G})), \quad (5.3b)$$

$$\mathbf{U}^{(l)}[t] = \sigma(\text{MP}_u^{(l)}([\mathbf{Z}^{(l)}[t] \parallel \mathbf{H}^{(l)}[t-1]], \mathcal{G})), \quad (5.3c)$$

$$\mathbf{C}^{(l)}[t] = \tanh(\text{MP}_c^{(l)}([\mathbf{Z}^{(l)}[t] \parallel \mathbf{R}^{(l)}[t] \odot \mathbf{H}^{(l)}[t-1]], \mathcal{G})), \quad (5.3d)$$

$$\mathbf{H}^{(l)}[t] = \mathbf{U}^{(l)}[t] \odot \mathbf{H}^{(l)}[t-1] + (\mathbf{1} - \mathbf{U}^{(l)}[t]) \odot \mathbf{C}^{(l)}[t]. \quad (5.3e)$$

The operators $\text{MP}_r^{(l)}(\cdot)$, $\text{MP}_u^{(l)}(\cdot)$, and $\text{MP}_c^{(l)}(\cdot)$ denote the MPNN layers for the reset, update, and candidate gates, respectively, and the activation functions $\sigma(\cdot)$ and $\tanh(\cdot)$ denote the sigmoid and hyperbolic tangent. This gated GRNN follows the general form of the graph GRU presented in [Sec. 2.3.2 \(2.55\)–\(2.58\)](#), with the key distinction that the message-passing operators here employ the anisotropic message function defined below rather than the diffusion or spectral convolutions discussed in [Sec. 2.3.2](#). The hidden state of the gated-GRNN at each layer l is initialised as a linear function of the node embeddings

$$\mathbf{H}^{(l)}[0] = \mathbf{V}\mathbf{W}_{\text{init}}^{(l)} + \mathbf{b}_{\text{init}}^{(l)}, \quad (5.4)$$

where $\mathbf{W}_{\text{init}}^{(l)} \in \mathbb{R}^{d_q \times d}$ and $\mathbf{b}_{\text{init}}^{(l)} \in \mathbb{R}^d$, which takes the different characteristics of each time series into account when initialising the state, thus reducing the need to rely on a long observation history [Montero-Manso and Hyndman, 2021].

Message-passing layers. The MPNN layers for each gate in the GRNN, denoted as MP, compute the hidden state of the i -th node as

$$\mathbf{h}_i^{(l+1)}[t] = \gamma^{(l)} \left(\mathbf{h}_i^{(l)}[t], \sum_{j \in \mathcal{N}_i} \rho^{(l)}(\mathbf{o}_i^{(l)}, \mathbf{o}_j^{(l)}, \mathbf{e}_{ij}) \right), \quad (5.5)$$

where $\mathbf{o}_i^{(l)} \in \mathbb{R}^{2d}$ represents the input of the gate at layer l , $\mathcal{N}_i$ denotes the set of nodes connected to i , $\rho^{(l)}(\cdot)$ is a message function, $\gamma^{(l)}(\cdot)$ is an update function, and the summation serves as a permutation invariant aggregation over neighbouring nodes. This instantiates the general MPNN framework of (2.42) with summation as the aggregation operator. Specifically, the message function $\rho^{(l)}(\cdot)$ is implemented as

$$\tilde{\mathbf{m}}_{ij}^{(l)} = \text{MLP}_{\text{msg}}^{(l)}([\mathbf{o}_i^{(l)} \parallel \mathbf{o}_j^{(l)}]), \quad (5.6a)$$

$$\alpha_{ij}^{(l)} = \sigma(\mathbf{W}_{\text{msg-}\alpha}^{(l)} \tilde{\mathbf{m}}_{ij}^{(l)}), \quad (5.6b)$$

$$\mathbf{m}_{ij}^{(l)} = \alpha_{ij}^{(l)} \tilde{\mathbf{m}}_{ij}^{(l)}, \quad (5.6c)$$

where the messages along non-zero edges $\mathbf{e}_{ij}$ are weighted according to an inferred scalar value $\alpha_{ij}^{(l)}$, which resides on the interval $[0, 1]$. The scalar value $\alpha_{ij}^{(l)}$ allows for anisotropic message passing, which assists in learning latent edge attributes such as the coupling strength between nodes. This approach provides flexibility in edge weighting while maintaining the fixed geometric structure, as opposed to methods that learn time-varying connectivity graphs [Jenkins, Varidhisi, et al., 2025]. Messages are then aggregated at node i using the summation, $\mathbf{m}_i^{(l)} = \sum_{j \in \mathcal{N}_i} \mathbf{m}_{ij}^{(l)}$. Finally, the update function $\gamma^{(l)}(\cdot)$ is implemented with a residual skip connection as

$$\mathbf{u}_i^{(l)} = \text{MLP}_{\text{update}}^{(l)}([\mathbf{h}_i^{(l)} \parallel \mathbf{m}_i^{(l)}]), \quad (5.7a)$$

$$\mathbf{h}_i^{(l+1)}[t] = \mathbf{u}_i^{(l)} + \mathbf{W}_{\text{skip}}^{(l)} \mathbf{o}_i^{(l)}, \quad (5.7b)$$

where $\mathbf{W}_{\text{skip}}^{(l)} \in \mathbb{R}^{d \times 2d}$ and $\mathbf{h}_i^{(l+1)} \in \mathbb{R}^d$. The parameters of the MPNN layers, MP, are defined separately for each gate and layer. This message-passing mechanism loosely resembles the gated graph convolution introduced in [Bresson and Laurent, 2017].

Iterative imputations. To learn effective state space representations while processing the data sequentially, missing values must be accounted for. To do this, first and second-stage imputations as in [Cini, Marisca, and Alippi, 2022] are performed iteratively within the spatiotemporal encoder

block. The first stage imputation performs one-step-ahead prediction via a linear readout as

$$\widehat{\mathbf{X}}_{\text{stage-1}}[t+1] = \mathbf{H}^{(L)}[t] \mathbf{W}_{\text{stage-1}} + \mathbf{b}_{\text{stage-1}}, \quad (5.8)$$

which is used in place of the missing values. The second stage imputation acts as a type of regularisation, estimating the value at node i using the previous hidden states, as well as the masks and first-stage imputations at the neighbouring nodes. The second stage imputation involves first computing an intermediate representation of each node, given by

$$\mathbf{h}'_i[t+1] = \gamma_{\text{stage-2}} \left(\mathbf{h}_i^{(L)}[t], \sum_{j \in \mathcal{N}_i} \rho_{\text{stage-2}}([\widehat{\mathbf{x}}_j^{(1)}[t+1] \parallel \mathbf{h}_j^{(L)}[t] \parallel \mathbf{m}_j[t+1]), \mathbf{e}_{ij}) \right), \quad (5.9)$$

where $\mathbf{h}'_i[t+1] \in \mathbb{R}^d$ and the message passing is implemented using a diffusion graph convolution [Atwood and Towsley, 2016]. Next, a linear readout layer performs the second stage imputation as

$$\widehat{\mathbf{X}}_{\text{stage-2}}[t+1] = [\mathbf{H}'[t+1] \parallel \mathbf{H}^{(L)}[t]] \mathbf{W}_{\text{stage-2}} + \mathbf{b}_{\text{stage-2}}, \quad (5.10)$$

which is used again in place of the missing values in $\mathbf{X}[t+1]$.

The process detailed so far in the spatiotemporal encoding block is repeated for $\mathbf{X}[t+1]$, to learn representations $\mathbf{H}'[t+2]$ and $\mathbf{H}^{(L)}[t+1]$, and so forth, until a set of representations $\mathbf{H}'[t:t+W]$ and $\mathbf{H}^{(L)}[t:t+W]$, and imputations $\widehat{\mathbf{X}}_{\text{stage-1}}[t:t+W]$ and $\widehat{\mathbf{X}}_{\text{stage-2}}[t:t+W]$ are calculated for the whole window length W . To summarise the spatiotemporal encoding block detailed in this section, the following shorthand notation is used

$$\begin{aligned} & \langle \mathbf{H}'[t:t+W], \mathbf{H}^{(L)}[t:t+W], \widehat{\mathbf{X}}_{\text{stage-1}}[t:t+W], \widehat{\mathbf{X}}_{\text{stage-2}}[t:t+W] \rangle \\ &= \text{ST-Encoder}(\mathbf{X}[t:t+W], \mathbf{M}[t:t+W], \mathbf{G}, \mathbf{V}). \end{aligned} \quad (5.11)$$

The Decoder

The sequences are encoded in both the forward and backward directions, to form

$$\begin{aligned} & \langle \mathbf{H}'_{\text{fwd}}, \mathbf{H}^{(L)}_{\text{fwd}}, \widehat{\mathbf{X}}_{\text{fwd,stage-1}}, \widehat{\mathbf{X}}_{\text{fwd,stage-2}} \rangle \\ &= \text{ST-Encoder}(\mathbf{X}[t:t+W], \mathbf{M}[t:t+W], \mathbf{G}, \mathbf{V}), \end{aligned} \quad (5.12)$$

and

$$\begin{aligned} & \langle \mathbf{H}'_{\text{bwd}}, \mathbf{H}^{(L)}_{\text{bwd}}, \widehat{\mathbf{X}}_{\text{bwd,stage-1}}, \widehat{\mathbf{X}}_{\text{bwd,stage-2}} \rangle \\ &= \text{ST-Encoder}(\mathbf{X}[t+W:t], \mathbf{M}[t+W:t], \mathbf{G}, \mathbf{V}), \end{aligned} \quad (5.13)$$

respectively, where $t+W:t$ denotes the time reversed/backward sequence. The bidirectional processing ensures that information from both past and future observations informs reconstruction at each time step, following the imputation-specific adaptation discussed in [Sec. 2.3.2](#). The decoder combines the hidden state representations from both directions to perform a final imputation at time τ ,

$$\begin{aligned} \hat{\mathbf{Y}}[\tau] = \text{MLP}_{\text{dec}}(&[\mathbf{H}'_{\text{fwd}}[\leq \tau] \parallel \mathbf{H}_{\text{fwd}}^{(L)}[\leq \tau] \parallel \mathbf{H}'_{\text{bwd}}[\geq \tau] \\ &\parallel \mathbf{H}_{\text{bwd}}^{(L)}[\geq \tau] \parallel \mathbf{M}[\tau] \parallel \mathbf{G} \parallel \mathbf{V}]), \end{aligned} \quad (5.14)$$

where the notation $[\leq \tau]$ refers to an index of the hidden states at time τ (which includes information from times $\leq \tau$) and similarly, $[\geq \tau]$ refers to the index at time τ (which includes information from times $\geq \tau$). The decoder takes a form similar to that in [\[Cini, Marisca, and Alippi, 2022\]](#) except patient-specific parameters are also provided to facilitate personalisation. The final imputed values for the entire window length are denoted $\hat{\mathbf{Y}}[t:t+W]$.

Computational Complexity

As a T&S GRNN, FIBMAP's computational complexity scales with both the temporal window length, W , and spatial dimensions, N . Following the framework described by Cini, Marisca, Zambon, et al. [\(2025\)](#), the time complexity scales as $\mathcal{O}(W(N + L|\mathcal{E}|))$, where L is the number of GRNN layers and $|\mathcal{E}|$ represents the maximum number of edges across time steps. The spatial resolution of the imputation map, N , fundamentally affects the computational cost. In a triangulated mesh representation of the atrial surface, the number of edges scales approximately linearly as $\mathcal{O}(N)$, giving an overall complexity of $\mathcal{O}(WNL)$. In practice, the computational complexity may further increase because higher spatial resolutions enlarge the graph diameter, necessitating larger temporal windows W to enable effective information propagation across the entire chamber surface. This creates a fundamental trade-off between spatial resolution and computational complexity, where finer discretisation leads to super-linear increases in computational requirements.

5.3.3 FibMap: Optimisation

To quantify the uncertainty of the reconstructed dynamics, FIBMAP is formulated as a quantile regression [\[Koenker and Hallock, 2001\]](#). In general, a quantile regression aims to estimate the conditional quantile function, $Q_Y(q|\mathbf{B})$, which represents the q -th quantile of the response variable, Y , given the predictor variables, $\mathbf{B}$. This is given by $Q_Y(q|\mathbf{B}) = \inf\{y \in \mathbb{R} \mid P(Y \leq y \mid \mathbf{B}) \geq q\}$, where $P(Y \leq y \mid \mathbf{B})$ is the conditional cumulative distribution function of Y given $\mathbf{B}$, and $\inf$ represents the infimum of the set.

In this chapter, conditional quantile functions for $q \in \mathcal{C}$, where $\mathcal{C} = \{q_1, q_2, \dots, q_{|\mathcal{C}|}\}$, are pre-

dicted for each of the imputed values. This is done by computing the following masked quantile/pinball loss function

$$\mathcal{L}(\hat{\mathbf{Y}}[t:t+W], \mathbf{Y}[t:t+W], \bar{\mathbf{M}}[t:t+W]) = \frac{\sum_{\tau=t}^{t+W} \sum_{i=1}^N \bar{m}_i[\tau] \mathcal{L}_{\tau,i}(\hat{\mathbf{y}}_i[\tau], y_i[\tau])}{\sum_{\tau=t}^{t+W} \sum_{i=1}^N \bar{m}_i[\tau]}, \quad (5.15)$$

where $\hat{\mathbf{Y}}[t] \in \mathbb{R}^{N \times |\mathcal{C}|}$, $\mathbf{Y}[t] \in \mathbb{R}^{N \times 1}$, and $\bar{\mathbf{M}}[t] \in \{0, 1\}^{N \times 1}$ represent the predicted values, target values, and the evaluation mask, respectively, at time t . The function $\mathcal{L}_{\tau,i}$ computes the average pinball loss at time τ and node i , which is given by

$$\mathcal{L}_{\tau,i}(\hat{\mathbf{y}}_i[\tau], y_i[\tau]) = \frac{1}{|\mathcal{C}|} \sum_{c=1}^{|\mathcal{C}|} (y_i[\tau] - \hat{\mathbf{y}}_i[\tau][c])(q_c - \mathbb{1}\{y_i[\tau] - \hat{\mathbf{y}}_i[\tau][c] < 0\}), \quad (5.16)$$

where c refers to the channel of $\hat{\mathbf{y}}_i[\tau]$ (prediction vector at time τ and node i) which contains the prediction of the c -th conditional quantile, $Q_y(q_c | \dots)$, and $\mathbb{1}\{y_i[\tau] - \hat{\mathbf{y}}_i[\tau][c] < 0\}$ represents the indicator function, equal to 1 if $y_i[\tau] - \hat{\mathbf{y}}_i[\tau][c] < 0$, and 0 otherwise. In practice, the decoder is modified to predict a value for each quantile (rather than a single value) and the average pinball loss is computed from each predicted quantile value as in (5.16).

For FIBMAP, the following loss function is minimised

$$\begin{aligned} \mathcal{L}_{\text{loss}} = & \mathcal{L}(\hat{\mathbf{Y}}[t:t+W], \mathbf{X}[t:t+W], \bar{\mathbf{M}}[t:t+W]) \\ & + \mathcal{L}(\hat{\mathbf{X}}_{\text{fwd,stage-1}}[t:t+W], \mathbf{X}[t:t+W], \bar{\mathbf{M}}[t:t+W]) \\ & + \mathcal{L}(\hat{\mathbf{X}}_{\text{fwd,stage-2}}[t:t+W], \mathbf{X}[t:t+W], \bar{\mathbf{M}}[t:t+W]) \\ & + \mathcal{L}(\hat{\mathbf{X}}_{\text{bwd,stage-1}}[t:t+W], \mathbf{X}[t:t+W], \bar{\mathbf{M}}[t:t+W]) \\ & + \mathcal{L}(\hat{\mathbf{X}}_{\text{bwd,stage-2}}[t:t+W], \mathbf{X}[t:t+W], \bar{\mathbf{M}}[t:t+W]), \end{aligned} \quad (5.17)$$

where $\mathcal{L}$ is given in (5.15) and each component of the loss is specific to a different imputation stage and processing direction. The specification of the evaluation mask, $\bar{\mathbf{M}}$, differs depending on training and fine-tuning as detailed in the following sections.

5.4 Experiments

5.4.1 Datasets

A fundamental challenge in developing and validating imputation mapping methods is obtaining high-fidelity ground truth data that captures the complete, simultaneous spatiotemporal dynamics of cardiac fibrillation, while remaining clinically relevant. While routine clinical contact mapping is limited to sequential localised measurements from a catheter sampling different regions

at different times, training and validating reconstruction algorithms requires a global ground truth of simultaneous observations across the entire chamber surface. This lack of a simultaneous, global ground truth within the clinical contact mapping workflow necessitates alternative validation strategies and data modalities. To address this requirement, two primary physiological datasets that provide suitable ground truth are utilised to benchmark FIBMAP’s ability to reconstruct a simultaneous global view from sparse, sequential input.

The first dataset consists of high-resolution ex vivo optical mapping of VF in 17 Sprague-Dawley rat hearts. As an imaging modality, optical mapping uses light-sensitive dyes to record electrical activity simultaneously across the entire heart surface (Sec. 3.5), providing an established reference for capturing complex surface dynamics [Efimov, Nikolski, and Salama, 2004]. This approach provides high-fidelity data of the heart’s electrical patterns, enabling rigorous validation of imputation mapping under controlled conditions. The use of VF as a surrogate is supported by evidence that the fundamental dynamics of fibrillation are mechanistically conserved across both cardiac chambers and mammalian species [Pandit and Jalife, 2013; Filgueiras-Rama and Jalife, 2016].

The second dataset comprises clinical non-contact mapping from 51 persistent AF patients recorded with the AcQMap system (Acutus Medical, Inc.) [Grace et al., 2019]. While non-contact mapping is the leading modality for recording simultaneous whole-atrial dynamics in human AF, its clinical use is limited by the inherent constraints of signal inverse-reconstruction (Sec. 3.5) [Rudy and Burnes, 1999; Roney et al., 2017]. Despite this, it is utilised here as a “silver-standard” ground truth for a proof-of-concept in humans. Together, these datasets enable the FIBMAP architecture to be validated across real samples of cardiac fibrillation dynamics, spanning different species, arrhythmia types, and recording modalities.

To further demonstrate practical translation, a secondary translational experiment is included using real-world contact mapping recordings from clinical procedures [Hong et al., 2019]. This supplemental study showcases how FIBMAP’s architecture can bridge the gap between sparse clinical measurements and global whole-atrial dynamics within existing clinical workflows. By using standard multipolar catheters, specifically the Advisor™ HD Grid catheter, this experiment demonstrates the potential for whole-atrial reconstruction without requiring specialised, simultaneous mapping hardware. The primary objective of this extension is to show that a pre-trained FIBMAP model can be successfully transferred to different hardware to achieve a physiologically reasonable reconstruction. The technical characteristics and processing of these datasets are detailed in the following subsections.

Rat VF Validation Dataset: High-Resolution *Ex Vivo* Optical Mapping

For model development and primary validation of the FIBMAP architecture against a gold standard, high-resolution optical mapping recordings of VF from 17 *ex vivo* Sprague-Dawley rat hearts with durations of 4 seconds were used. Optical mapping records fluorescence signals proportional to action potentials using a 128×80 -pixel camera (≈ 0.2 mm pixel spacing) at 1000 Hz, providing high-resolution recordings of chaotic fibrillation dynamics. This dataset yielded approximately 68,000 temporal frames captured across 174,000 spatial pixels simultaneously (totalling over 11.8 billion unique spatiotemporal data points). While VF differs from AF in chamber and substrate, both exhibit complex spatiotemporal propagation patterns (multiple wavefronts, rotational activity, chaotic dynamics), making VF recordings valuable for validating the core methodology.

All animal procedures were performed in accordance with the UK Animals (Scientific Procedures) Act 1986, ARRIVE guidelines, and Directive 2010/63/EU, and were approved by the Imperial College London Ethical Review Board (project licences PEE7C76CD and PCA5EE967). Rats (9–12 weeks old, 250–300 g) were anaesthetised with 5% isoflurane (95% oxygen mix) by inhalation in an induction chamber and euthanised via cervical dislocation. Hearts were rapidly explanted and Langendorff-perfused following established protocols [Handa, X. Li, et al., 2021].

Raw optical signals underwent established preprocessing [Handa, X. Li, et al., 2021; Laughner, F. S. Ng, et al., 2012; Umapathy et al., 2010]: spatial and temporal filtering, baseline drift removal, and min-max normalisation to emphasise propagation patterns and ensure equal training contribution across spatial locations. Signals were downsampled to 42×26 -pixel resolution (~ 0.6 mm spacing) at 200 Hz to preserve fine-scale fibrillatory dynamics while maintaining computational tractability. This yielded a processed dataset of approximately 13,600 total temporal frames across 18,600 unique spatial pixels (totalling nearly 253 million spatiotemporal data points). For each heart, a graph was constructed by connecting each pixel (node) to its 8 nearest neighbours (neighbouring pixels on the horizontal, vertical, and diagonal) to approximate the ventricular surface connectivity.

Clinical Validation Dataset: Non-Contact Mapping of Persistent AF with AcQMap

For model development and primary validation of the FIBMAP architecture on human AF, clinical non-contact recordings from 51 persistent AF patients with durations of 5–20 seconds, obtained using the AcQMap system (Acutus Medical, Inc.) prior to ablation, were used. This dataset is comprised of over 2.3 million temporal frames across 183,000 spatial nodes (433 billion unique spatiotemporal data points). The AcQMap system uses non-contact electrodes within the atrial

Table 5.1: Characteristics of patients in the AcQMap dataset. Values are mean $\pm$ SD or n (%). The asterisk denotes the presence of missing values; the number of missing values is shown as n^* .

Total ($N = 51$)	
<i>Demographics</i>	
Mean age (years)	64 ± 11
Male	35 (69%)
Mean BMI	29 ± 5
<i>Comorbidities</i>	
Hypertension	20 (39%)
Diabetes mellitus	5 (10%)
Heart failure	17 (33%)
CHA ₂ DS ₂ -VASc score >2	14 (27%)
<i>Medications</i>	
Antiarrhythmics: Beta-blockers	33* (87%, $n^*=13$)
Antiarrhythmics: Amiodarone	11* (29%, $n^*=13$)
Anticoagulants	38* (100%, $n^*=13$)
Statins	16* (42%, $n^*=13$)

cavity to sense intra-cavitary unipolar electrograms at 3000 Hz, from which dipole density measurements (charge density in Coulombs/cm) are inversely derived to reconstruct electrical activity across approximately 3500 nodes on the atrial surface simultaneously. This inverse reconstruction approach provides whole-atrial simultaneous coverage without direct tissue contact, though with acknowledged spatial resolution limitations due to the ill-posed nature of the inverse problem and necessary regularisation [Rudy, 2017; Roney et al., 2017]. Despite these limitations and the system’s limited clinical adoption [Narayan, Krummen, et al., 2012; Buch et al., 2016; Roney et al., 2017], these recordings provide realistic samples of human AF dynamics for assessing FIBMAP’s feasibility on clinical data.

Patients were treated between 2016 and 2023 at two centres in the United Kingdom: The Royal Brompton Hospital, London, and the John Radcliffe Hospital, Oxford. Patient demographics and comorbidities are detailed in Tab. 5.1. Procedures involved left atrial mapping, pulmonary vein isolation, and in some cases adjunctive linear ablations and cardioversion. All procedures were performed with informed consent and approved by the London-Surrey Research Ethics Committee (REC: 20/LO/0150), conducted per the Declaration of Helsinki, and sponsored by Oxford University Hospitals NHS Foundation Trust (IRAS: 255290).

Raw dipole density signals were preprocessed: spatially resampled to 500 nodes via k -means clustering, temporally filtered and resampled to 200 Hz, and min-max normalised at the node level

to emphasise propagation patterns. This yielded a processed dataset of approximately 158,000 total temporal frames across 25,500 unique spatial node locations (totalling over 4.0 billion unique spatiotemporal data points). This resolution preserves clinically relevant AF dynamics (phase singularities, wavefront propagation) while ensuring computational feasibility. For each patient, a graph was constructed by triangulating the discretised atrial surface [Bernardini et al., 1999].

Supplemental Dataset: Routine Sequential Contact Mapping with the HD Grid

To showcase the translation of FIBMAP to routine clinical workflows, a supplemental dataset of sequential contact mapping from three patients (under the same ethical approval as [Sec. 5.4.1](#)) was included, using the Advisor™ HD Grid catheter (16 electrodes, 4×4 grid, 3 mm spacing) and EnSite Precision system (Abbott). For these select patients, both HD Grid and AcQMap recordings were collected non-contemporaneously before ablation. This dual-modality approach provides a unique opportunity to compare FIBMAP’s reconstructions (from HD Grid) against a silver-standard ground truth (AcQMap) within the same patient, serving as an initial assessment of whether the model captures patient-specific dynamical signatures despite the lack of direct temporal alignment.

HD Grid recordings (2035 Hz, 13–20 minutes) yielded 1.6–2.4 million temporal frames across 1100–4600 spatial points per patient. Recordings were then preprocessed as follows: spatial interpolation to the atrial surface, identification of active recording periods, voltage normalisation, spatial resampling to 500 nodes, and temporal resampling to 200 Hz. Graphs were constructed by triangulating the atrial surface [Bernardini et al., 1999].

5.4.2 Experimental Setup

Dataset Partitioning

The two primary benchmarking datasets (optical and AcQMap) were each partitioned at two levels: first across hearts/patients, then temporally within each heart.

At the heart level, datasets were split into training (70%), validation (10%), and test (20%) subsets, stratified to ensure that each subset contained a representative spectrum of fibrillation dynamics. To achieve this, Shannon entropy was first computed for the resampled time series at each node, and the cumulative distribution function of entropy across the chamber surface was constructed for each heart, serving as a summary of the spatiotemporal complexity of that recording. Pairwise similarity between hearts was then assessed via the Kolmogorov–Smirnov distance between their entropy distributions, where a smaller distance indicates more similar spatiotemporal dynamics.

Groups of dynamically similar hearts were identified by clustering a Laplacian eigenmap [Belkin and Niyogi, 2003] of the resulting distance matrix using k -means, with the number of clusters determined by the elbow method. The Laplacian eigenmap was constructed by forming a weighted adjacency matrix from a radial basis function kernel applied to the reciprocal of the Kolmogorov–Smirnov distances, followed by thresholding. Training, validation, and test sets were then formed by stratified sampling across these clusters, maintaining a consistent distribution of fibrillation dynamics within each subset.

At the temporal level, training and validation hearts had their time series split sequentially into training (85%), validation (5%), and test (10%) segments for model optimisation and selection, while held-out test hearts retained complete time series for fine-tuning and evaluation.

Model Training and Validation

Two separate FIBMAP models were trained: one for rat VF using optical mapping data and one for human AF using AcQMap data. To train FIBMAP on datasets with complete spatiotemporal ground truth while preparing it for sparse clinical data, a sequential contact mapping strategy was simulated. Specifically, data collection with a multipolar catheter was simulated by a spatial patch of observations that moves sequentially across the chamber surface, with configurable parameters including patch size (catheter surface area), dwell time (recording duration at each location), and spatial overlap between successive positions.

For optical mapping, catheter patches were simulated as rectangular pixel regions traversing the ventricular surface, with sizes ranging from 3×3 to 15×15 pixels (1–28% surface coverage), stride patterns varying from non-overlapping to 75% overlap between successive patches, and dwell times from 3–150 frames (20–760 ms at 200 Hz sampling). The sequential patching algorithm calculated appropriate padding to ensure complete surface coverage and systematically traversed the heart surface, with each patch position validated to contain cardiac tissue before inclusion in the observation mask. For the AcQMap dataset, patches were defined using k -means clustering of the mesh vertices, with catheter surface areas ranging from 2.5–50% of the total atrial surface, dwell times from 0.2–4 seconds, and spatial overlap from 0–3 additional clusters shared between adjacent patches. Catheter movement was simulated as a self-avoiding random walk: an initial patch was randomly selected, and remaining patches were sampled without replacement with probability proportional to the distance from the current patch, mimicking realistic navigation patterns. For each simulated mapping procedure, the sequence of patches was repeated cyclically until the full recording duration was covered.

Table 5.2: Hyperparameter values tested during FIBMAP training.

Hyperparameter	Range tested
Batch size, B	$\{16, 32, 64\}$
Input window length, W (timesteps)	$\{20, 30, 40, 50\}$
Hidden size, d	$\{64, 128, 256\}$
Node embedding size, q	$\{16, 32, 64\}$
Patient embedding size, r	$\{16, 32, 64\}$
Layers, L	$\{1, 2, 3\}$
Hidden layer neurons in MLP_{enc} and MLP_{dec}	$\{128, 256, 512, 1024\}$

Training employed a self-supervised learning approach in which these catheter parameters were augmented across training examples, reflecting the diversity of clinical practice where catheter types, operator preferences, and anatomical constraints create heterogeneous data acquisition scenarios. The model optimised a whole-surface reconstruction loss where all spatial locations and time points contributed regardless of whether they were observed or unobserved. This training objective forced the network to propagate information from observed patches across unobserved regions, learning universal fibrillation propagation patterns through shared parameters while capturing patient-specific characteristics through embedding parameters. The hyperparameter search space is detailed in [Tab. 5.2](#).

Deployment via Fine-Tuning

To simulate clinical deployment where only sparse contact measurements are available, the trained model was adapted to new patients in the test set through fine-tuning. This procedure preserves the knowledge for whole-surface reconstruction acquired during training by fixing the shared model parameters, while quickly personalising the model to new patients by optimising only the node and patient embedding parameters using an observed-patch reconstruction loss ([Fig. 5.2](#)). The observed-patch loss restricts the computation in [\(5.15\)](#) to only the spatiotemporal locations where measurements are available, enabling rapid patient-specific adaptation from sparse observations.

During fine-tuning, sequential contact mapping was simulated using fixed catheter parameters to ensure consistent evaluation. For the optical mapping data, 9×9 pixel patches ($\sim 10\%$ instantaneous surface coverage) with 10-frame (50 ms) dwell times and non-overlapping placement were used. This short dwell time was selected to accommodate the significantly shorter cycle lengths characteristic of rat VF. For AcQMap, patches were simulated to represent $\sim 10\%$ instantaneous atrial surface area with 1-second dwell times and no spatial overlap, selected to approximate the sequential contact mapping process used clinically. In both cases, a fixed self-avoiding walk de-

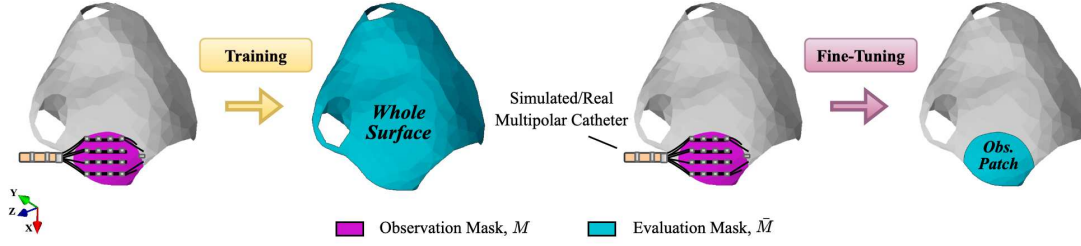


Figure 5.2: Training and fine-tuning workflow for FIBMAP. During training (left), FIBMAP learns from complete ground truth recordings with simulated sequential contact mapping (observation mask, $\mathbf{M}$, shown in magenta). The model is optimised using a whole-surface reconstruction loss (cyan) computed across all spatial locations and time points, forcing the network to propagate information from observed patches to unobserved regions. During fine-tuning (right), only patient-specific embedding parameters are optimised using an observed-patch reconstruction loss (cyan) restricted to sparse measurements from simulated or real multipolar catheter recordings, enabling rapid patient-specific adaptation whilst shared parameters remain fixed.

terminated the catheter trajectory.

Fine-tuning hyperparameters were determined on validation hearts, which retain ground-truth whole-surface signals. A random hyperparameter search was conducted across learning rates in $[0.0005, 0.005]$ and batch sizes in $\{16, 32, 64, 128\}$, with each configuration run for 100 epochs on the AcQMap dataset and 25 epochs on the optical mapping dataset, using early stopping based on whole-surface mean absolute error (MAE) on the validation temporal segment. The best-performing configuration was selected by minimising whole-surface MAE on the test temporal segment, yielding a learning rate of 0.005 and a batch size of 16. For both datasets, a strong correlation (Pearson’s $r \geq 0.97$, $p < 0.05$) was observed between observed-patch and whole-surface reconstruction losses (Fig. 5.3), confirming that improvements in patch-level optimisation translated to accurate whole-surface reconstruction without overfitting.

For final evaluation on held-out test hearts, the pre-trained FIBMAP model was fine-tuned using the optimal configuration for 100 epochs with early stopping on the observed-patch loss. Unlike the validation procedure, all time steps were used for fine-tuning (no sequential temporal split applied).

Baseline Methods

FIBMAP’s performance on the optical mapping and AcQMap test sets was benchmarked against five baseline imputation methods representing different modelling paradigms. Mean imputation replaces missing values with the spatial average across observed nodes at each time step. MF performs low-rank approximation (rank 10) to capture dominant spatiotemporal patterns. Both are transductive methods computed directly from each test patient’s available data without requiring

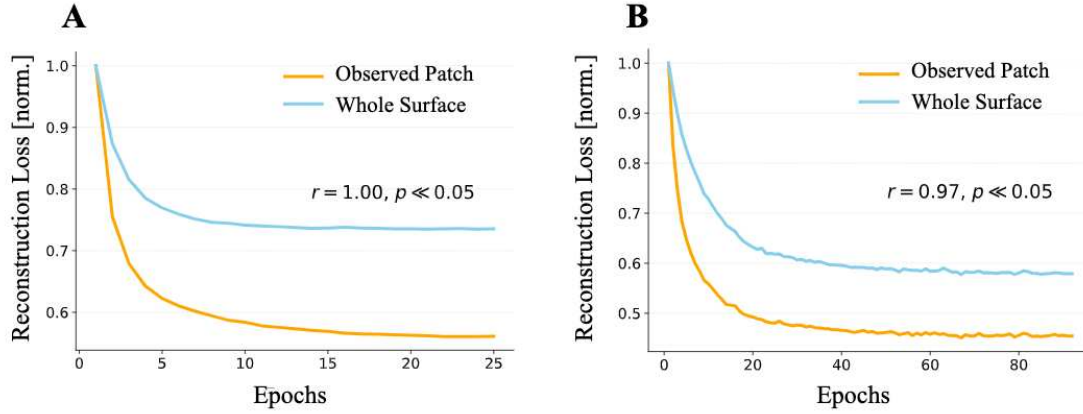


Figure 5.3: Fine-tuning convergence on validation hearts/patients. Reconstruction loss curves during fine-tuning for A) optical mapping (rat VF) and B) AcQMap (human AF) validation datasets, showing observed-patch loss (orange) and whole-surface loss (blue). Strong correlation between loss curves (Pearson's $r \geq 0.97$, $p < 0.05$) confirms that improvements in patch-level fitting translate to accurate whole-surface reconstruction without overfitting. Both datasets show rapid loss plateauing, with most improvement achieved within early epochs.

a training phase. To assess the role of temporal context, univariate RNNs and bidirectional RNNs (Bi-RNN) that process each node's time series independently, ignoring spatial structure, were implemented. Finally, a TTS-Transformer baseline was included, applying temporal self-attention followed by spatial self-attention with 8 heads [Vaswani et al., 2017].

The RNN, Bi-RNN, and TTS-Transformer models were pre-trained using identical training protocols to FIBMAP (the same dataset splits, hyperparameter configurations, and catheter simulation procedures), but optimised using MAE loss rather than the quantile loss of (5.15). To assess the benefit of patient-specific adaptation, fine-tuned variants (RNN+FT, Bi-RNN+FT, TTS-Transformer+FT) were created by adapting pre-trained baselines for 50 epochs per test patient using the same observed-patch reconstruction loss as FIBMAP. Fine-tuning optimised selected layers to enable adaptation while preserving learned representations: input-to-hidden weights and readout layers for RNN models, and input encoder, readout, and normalisation layers for the TTS-Transformer.

Evaluation Metrics

All metrics were computed over unobserved regions only, and all models were evaluated across 5 random seeds with mean and standard deviation reported. Reconstruction fidelity was assessed using MAE, MSE, and mean relative error (MRE).

Beyond pointwise accuracy, the ability of imputation models to reconstruct the characteristic dynamical features of cardiac fibrillation was evaluated by detecting phase singularity (PS) locations

Table 5.3: Inter-observer agreement for phase singularity detection on human AF AcQMap recordings and comparison with automated 3D topological charge algorithm [Rantner et al., 2007]. Manual annotations by two independent trained observers show excellent agreement at high precision threshold (~ 0.5 cm, 0.1 s), substantially outperforming automated detection against both observers. Statistics computed across all test patients with manually annotated recordings.

	F1-Score	Cohen's κ
Inter-Observer Agreement	0.827 ± 0.243	0.885 ± 0.187
Automated vs Observer 1	0.238 ± 0.130	0.352 ± 0.148
Automated vs Observer 2	0.292 ± 0.151	0.392 ± 0.132

across space and time. PS locations represent the topological vortex cores around which rotational waves revolve (Sec. 3.3). PS detection accuracy was quantified using F1-scores computed across four spatiotemporal precision thresholds: High (~ 0.6 mm, 15 ms for optical; ~ 0.5 cm, 0.1 s for AcQMap), Moderate (~ 1.2 mm, 30 ms; ~ 1.0 cm, 0.2 s), Standard (~ 1.8 mm, 45 ms; ~ 1.5 cm, 0.3 s), and Lenient (~ 2.4 mm, 60 ms; ~ 2.0 cm, 0.4 s). For optical mapping, PS locations were automatically detected using convolutional topological charge algorithms [Bray et al., 2001]. For AcQMap, the mesh-based topological charge algorithm [Rantner et al., 2007] proved unreliable at 500-node resolution, necessitating manual annotation. Two independent trained observers manually annotated PS locations in the central 100 frames (0.5 s) of each test recording using a custom graphical user interface. The 100-frame window balanced comprehensive evaluation with feasible annotation time, as manual annotation of full-length recordings would have been prohibitively time-consuming. Inter-observer agreement was excellent (see Tab. 5.3; F1: 0.827 ± 0.243 , Cohen's κ : 0.885 ± 0.187 at High Precision threshold), substantially outperforming automated detection (F1: 0.238 and 0.292). F1-score statistics were computed from 1000 bootstrap samples per model.

For the AcQMap test set, three of ten patients were excluded because their recording duration was insufficient for the simulated catheter to visit each spatial location at least once under the specified mapping parameters (10% surface area patches, 1-second dwell time, no spatial overlap), yielding results from seven test patients.

Supplemental Study: Showcase of Clinical Translation to HD Grid

To demonstrate the practical translation of a pre-trained FIBMAP model to routine clinical data, the human AF model from the AcQMap validation study was transferred to real EnSite Precision HD Grid contact mapping recordings via the fine-tuning procedure described in Tab. 5.4.2. Three persistent AF patients from the AcQMap test set had both HD Grid and AcQMap recordings collected non-contemporaneously before ablation. AcQMap recordings were 20 seconds in duration, while HD Grid recordings were significantly longer (13, 18, and 20 minutes per patient). The HD

Grid data consists of sequential contact mapping recordings from 16 electrodes arranged in a 4×4 grid array (3 mm spacing), along with the roving three-dimensional coordinates of each electrode within the atrium.

HD Grid preprocessing. To ensure compatibility with FIBMAP pre-trained on the AcQMap dataset, HD Grid recordings underwent several preprocessing steps. Sparse electrogram recordings were first mapped to a uniform discretisation of the atrial surface using nearest-neighbour interpolation (3 mm radius) to interpolate signals between electrodes. An observation mask identified active recording periods, after which signals were normalised using the maximum peak-to-peak voltage at each node. The data was then spatially resampled to 500 nodes using k -means clustering, low-pass filtered at 90 Hz, and temporally resampled to 200 Hz. Final min-max normalisation ensured all signals fell within $[0, 1]$. From these processed measurements, imputation maps were generated using the fine-tuning procedure of [Tab. 5.4.2](#), with the observed-patch reconstruction loss monitored for early stopping.

Cross-correlation methodology. As HD Grid and AcQMap recordings were collected minutes apart and AF exhibits significant beat-to-beat variability, direct point-by-point temporal alignment was not feasible. Clinical translation was therefore assessed by quantifying the statistical similarity between FIBMAP-imputed phase maps (reconstructed from sparse HD Grid observations) and the corresponding AcQMap recordings using a sliding-window cross-correlation analysis (see [Fig. 5.8A](#)).

Spatial alignment between AcQMap and FIBMAP imputation maps was achieved by centring the vertices of both geometries around the origin, applying rigid registration using the iterative closest point algorithm [Besl and McKay, 1992], and projecting data between geometries using $k=5$ nearest-neighbours regression. This enabled direct spatial comparison of phase dynamics across the two recording modalities despite differences in chamber geometry reconstruction.

Sliding-window cross-correlation was computed using window lengths varying from 0.5 to 4.0 seconds with a constant stride of 0.1 seconds. Pearson correlation coefficients were computed between the Hilbert phase transforms of the processed signals across all nodes and time points within each window pair, generating a cross-correlation matrix characterising the spatiotemporal similarity between recordings. For the primary analysis (0.5 s windows), this approach generated 197–198 windows from each AcQMap recording and 1800–6491 windows from each HD Grid recording, yielding 3.5×10^5 to 1.3×10^6 cross-correlation values per patient pair.

Baseline comparisons. Three types of correlations were computed: (i) intra-patient comparisons between AcQMap and FIBMAP imputation maps from the same patient, (ii) inter-patient comparisons between AcQMap from one patient and FIBMAP from a different patient, and (iii) shuffled baseline comparisons where both spatial (node) and temporal (frame) dimensions of the FIBMAP imputation maps were randomly permuted before computing intra-patient correlations with AcQMap. The inter-patient comparison establishes a baseline for correlations expected from different individuals, while the shuffled baseline establishes correlations expected by random chance. For each patient and window duration, the 99th percentiles of the intra-patient, inter-patient, and shuffled cross-correlation distributions were computed to characterise the upper tail of the correlation distributions. Statistical significance was assessed via bootstrap sampling (1000 rounds of 10 000 resamples per condition), enabling comparison of whether the 99th percentiles differed significantly across the three conditions at each window duration.

Notably, the lack of direct temporal alignment necessitates that the expected magnitude of absolute cross-correlations remains modest. The primary focus of this supplemental assessment is the statistical significance and relative strength of intra-patient similarity compared to the baselines. This approach evaluates whether FIBMAP captures stable, patient-specific dynamical signatures (reflected by significantly higher intra-patient cross-correlations) that persist across different recording modalities even as instantaneous beat-to-beat dynamics evolve. Further considerations regarding the refinement of this translational approach and the requirements for future large-scale prospective validation are detailed in [Sec. 5.6](#).

5.5 Results

The performance of the FIBMAP framework is evaluated through a multi-stage approach. First, a core proof-of-concept validation is established using two primary benchmarking datasets: foundational evaluation in *ex vivo* rat VF ($n = 17$; [Sec. 5.5.1](#)) and a human persistent AF cohort ($n = 51$; [Sec. 5.5.2](#)). Following this core validation, a supplemental translational showcase ([Sec. 5.5.3](#)) demonstrates the extension of FIBMAP to routine sequential contact mapping in standard clinical workflows.

5.5.1 Rat VF: Optical Mapping Validation

On the optical test set, FIBMAP accurately reconstructed complete ventricular activity from simulated sparse sequential contact mapping observations, with reconstruction losses plateauing rapidly during fine-tuning ([Fig. 5.3A](#)). [Fig. 5.4](#) shows representative reconstructions for three test hearts.

Table 5.4: Quantitative reconstruction performance on rat VF optical mapping test set. Models evaluated using simulated sequential contact mapping (9×9 pixel patches, $\sim 10\%$ instantaneous surface coverage, 50 ms dwell time, no spatial overlap). Reconstruction metrics: MAE, MSE, and MRE. PS detection evaluated using F1-score at standard precision threshold (~ 1.8 mm, 45 ms). Models include transductive baselines, pre-trained inductive baselines, and fine-tuned variants (FT). Best performing method in **bold**, second best underlined. Results reported as mean $\pm$ standard deviation across 5 random seeds.

Model	MAE $\downarrow$	MSE $\downarrow$	MRE (%) $\downarrow$	PS F1 $\uparrow$
<i>Transductive</i>				
Mean	0.2141 $\pm$ 0.0000	0.0717 $\pm$ 0.0000	44.94 $\pm$ 0.00	0.004 $\pm$ 0.000
MF	0.2164 $\pm$ 0.0005	0.0733 $\pm$ 0.0002	45.40 $\pm$ 0.10	0.351 $\pm$ 0.012
<i>Inductive</i>				
RNN	0.2059 $\pm$ 0.0000	0.0587 $\pm$ 0.0000	43.16 $\pm$ 0.00	0.591 $\pm$ 0.000
Bi-RNN	0.1874 $\pm$ 0.0000	0.0503 $\pm$ 0.0000	39.26 $\pm$ 0.00	<u>0.645$\pm$0.000</u>
TTS-Transformer	0.2022 $\pm$ 0.0000	0.0589 $\pm$ 0.0000	42.38 $\pm$ 0.00	0.613 $\pm$ 0.000
<i>Fine-tuned (FT)</i>				
RNN + FT	0.2051 $\pm$ 0.0001	0.0586 $\pm$ 0.0000	43.00 $\pm$ 0.01	0.605 $\pm$ 0.002
Bi-RNN + FT	<u>0.1766$\pm$0.0007</u>	<u>0.0476$\pm$0.0003</u>	<u>37.03$\pm$0.13</u>	0.609 $\pm$ 0.002
TTS-Transformer + FT	0.1992 $\pm$ 0.0019	0.0626 $\pm$ 0.0013	41.71 $\pm$ 0.39	0.573 $\pm$ 0.008
FIBMAP	0.0828$\pm$0.0004	0.0142$\pm$0.0002	17.41$\pm$0.07	0.795$\pm$0.046

From 10% instantaneous surface coverage, FIBMAP reconstructed full spatiotemporal activity, capturing multiple wavefronts and rotational activity invisible in partial observations. Reconstructed phase maps (Fig. 5.4A) revealed the characteristic chaotic VF dynamics, with FIBMAP preserving both spatial propagation patterns and temporal dynamics across the entire chamber. FIBMAP’s signal reconstructions at the pixel level demonstrated high fidelity to ground truth (Fig. 5.4B), maintaining accurate amplitude and phase even in unobserved regions. Prediction intervals (shaded regions, 10th–90th quantiles) widened when phase coherence was lost (circled region in Fig. 5.4B), demonstrating the model’s ability to quantify reconstruction confidence.

Quantitatively, FIBMAP significantly outperformed all baseline imputation methods (Tab. 5.4). FIBMAP achieved an MAE of 0.0828 ± 0.0004 , representing a 53% improvement over the next best baseline (Bi-RNN + FT: 0.1766 ± 0.0007). For PS detection at standard precision (~ 1.8 mm, 45 ms), FIBMAP achieved an F1-score of 0.795 ± 0.046 , a 23% improvement over the next-best method (Bi-RNN: 0.645 ± 0.000). Performance remained superior across all precision thresholds, from high (~ 0.6 mm, 15 ms: $F1 = 0.461 \pm 0.019$) to lenient (~ 2.4 mm, 60 ms: $F1 = 0.843 \pm 0.053$) (Tab. 5.5). In contrast, the next best baseline method (Bi-RNN + FT) failed to capture the complex spatiotemporal VF dynamics, producing phase maps with substantial errors (Fig. 5.4A) and poor signal reconstruction (Fig. 5.4B).

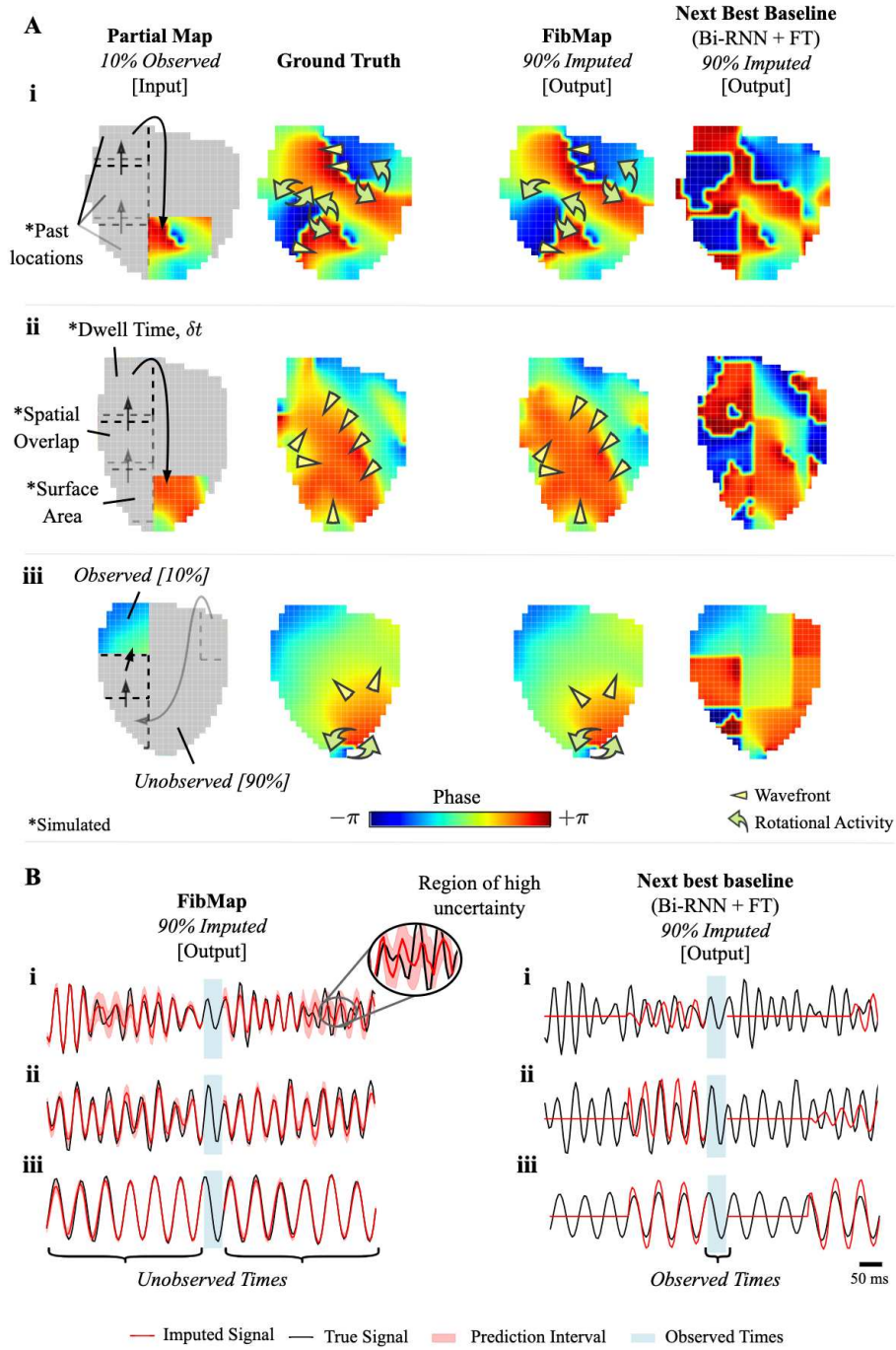


Figure 5.4: Qualitative reconstruction of rat ventricular fibrillation from simulated sparse sequential contact mapping. A) Representative phase map reconstructions for three test hearts (i–iii) showing partial observations (left, $\sim 10\%$ instantaneous surface coverage), ground truth optical mapping (centre-left), FIBMAP reconstructions (centre-right), and next-best baseline Bi-RNN+FT (right). FIBMAP accurately reconstructs complete spatiotemporal dynamics from sparse observations, capturing complex fibrillation features including multiple wavefronts (yellow arrows) and rotational activity (green arrows) that remain invisible in partial observations. Simulated catheter parameters: $\sim 10\%$ instantaneous surface area, no spatial overlap, 50 ms dwell time. B) Pixel-level signal reconstructions at representative nodes for the three test hearts, showing FIBMAP predictions (red lines), ground truth (black lines), prediction intervals (pink shaded regions, 10th–90th quantiles), and observed time periods (blue shaded regions). FIBMAP maintains high signal fidelity in unobserved regions, with prediction intervals appropriately widening during regions of high uncertainty (circled for heart i), demonstrating effective uncertainty quantification. The next-best baseline (Bi-RNN+FT, right panels) shows poor reconstruction fidelity with incorrect amplitude and phase relationships during unobserved periods.

Table 5.5: Phase singularity detection performance on rat VF optical mapping test set across precision thresholds. Models evaluated using simulated sequential contact mapping (9×9 pixel patches, $\sim 10\%$ surface coverage, 50 ms dwell time, no spatial overlap). F1-scores reported at four spatiotemporal precision thresholds. Best performing method in **bold**, second best underlined. Results reported as mean $\pm$ standard deviation across 5 random seeds.

Method	High Precision (≈ 0.6 mm, 15 ms)	Moderate Precision (≈ 1.2 mm, 30 ms)	Standard Precision (≈ 1.8 mm, 45 ms)	Lenient Precision (≈ 2.4 mm, 60 ms)
<i>Transductive</i>				
Mean	0.001 $\pm$ 0.000	0.002 $\pm$ 0.000	0.004 $\pm$ 0.000	0.004 $\pm$ 0.000
MF	0.096 $\pm$ 0.010	0.218 $\pm$ 0.014	0.351 $\pm$ 0.012	0.467 $\pm$ 0.013
<i>Inductive</i>				
RNN	0.253 $\pm$ 0.000	0.460 $\pm$ 0.000	0.591 $\pm$ 0.000	0.658 $\pm$ 0.000
Bi-RNN	<u>0.315$\pm$0.000</u>	<u>0.531$\pm$0.000</u>	<u>0.645$\pm$0.000</u>	<u>0.701$\pm$0.000</u>
TTS-Transformer	0.245 $\pm$ 0.000	0.477 $\pm$ 0.000	0.613 $\pm$ 0.000	0.686 $\pm$ 0.000
<i>Fine-tuned</i>				
RNN + FT	0.273 $\pm$ 0.001	0.487 $\pm$ 0.001	0.605 $\pm$ 0.002	0.671 $\pm$ 0.005
Bi-RNN + FT	0.297 $\pm$ 0.002	0.495 $\pm$ 0.002	0.609 $\pm$ 0.002	0.689 $\pm$ 0.003
TTS-Transformer + FT	0.242 $\pm$ 0.006	0.441 $\pm$ 0.008	0.573 $\pm$ 0.008	0.670 $\pm$ 0.004
FIBMAP	0.461$\pm$0.019	0.692$\pm$0.039	0.795$\pm$0.046	0.843$\pm$0.053

5.5.2 Human AF: Non-Contact Mapping Validation

Evaluation on the AcQMap test set demonstrated accurate reconstruction of complete atrial activity from simulated sparse sequential observations, with rapid convergence during fine-tuning (Fig. 5.3B). Fig. 5.5 shows representative reconstructions for three test patients. From 10% instantaneous surface coverage via simulated catheter patches, FIBMAP reconstructed full atrial activity, capturing multiple wavefronts and rotational activity invisible in partial observations. Reconstructed phase maps (Fig. 5.5A) revealed the characteristic chaotic dynamics of AF, with FIBMAP preserving both spatial propagation patterns and temporal dynamics across the entire chamber. FIBMAP’s signal reconstructions at individual tissue regions demonstrated high fidelity to ground truth (Fig. 5.5B), maintaining accurate amplitude and phase even in unobserved regions. Prediction intervals (shaded regions, 10th–90th quantiles) widened when amplitude variations occurred (circled region in Fig. 5.5B), demonstrating the model’s ability to quantify reconstruction confidence.

Quantitatively, FIBMAP significantly outperformed all baseline imputation methods (Tab. 5.6). FIBMAP achieved an MAE of 0.0706 ± 0.0007 , representing a 47% improvement over the next best baseline (MF: 0.1321 ± 0.0008). For PS detection at standard precision (~ 1.5 cm, 0.3 s), FIBMAP achieved an F1-score of 0.922 ± 0.016 , a 109% improvement over the next-best method (MF: 0.441 ± 0.100). Performance remained superior across all precision thresholds, from high (~ 0.5 cm, 0.1 s: $F1 = 0.487 \pm 0.059$) to lenient (~ 2.0 cm, 0.4 s: $F1 = 0.977 \pm 0.008$) (Tab. 5.7).

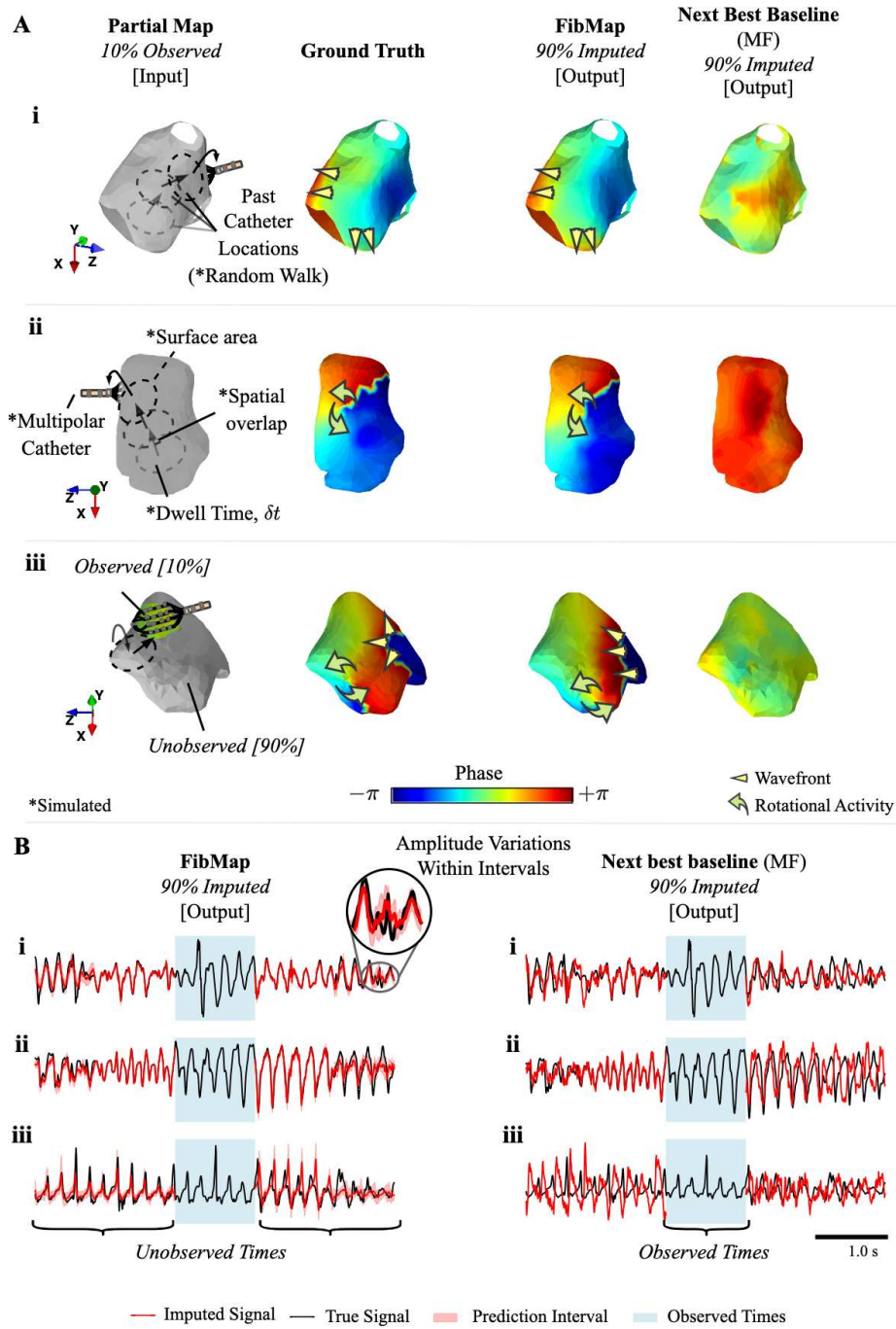


Figure 5.5: Qualitative reconstruction of human AF from simulated sparse sequential contact mapping. A) Representative phase map reconstructions for three test patients (i–iii) showing partial observations (left, $\sim 10\%$ instantaneous atrial surface coverage), ground truth AcQMap recordings (centre-left), FIBMAP reconstructions (centre-right), and next-best baseline MF (right). FIBMAP accurately reconstructs complete spatiotemporal AF dynamics from sparse observations, capturing complex fibrillation features including multiple wavefronts (yellow arrows) and rotational activity (green arrows) that remain invisible in partial observations. Simulated catheter parameters: $\sim 10\%$ instantaneous atrial surface area, no spatial overlap, 1 second dwell time. B) Node-level signal reconstructions at representative atrial locations for the three test patients, showing FIBMAP predictions (red lines), ground truth (black lines), prediction intervals (pink shaded regions, 10th–90th quantiles), and observed time periods (shaded blue regions). The next-best baseline (MF, right panels) shows poor reconstruction fidelity with loss of temporal dynamics and phase relationships during unobserved periods.

Table 5.6: Quantitative reconstruction performance on human AF AcQMap test set. Models evaluated using simulated sequential contact mapping (patches covering $\sim 10\%$ instantaneous atrial surface area, 1 s dwell time, no spatial overlap). Reconstruction metrics: MAE, MSE, and MRE. PS detection evaluated using F1-score at standard precision threshold (~ 1.5 cm, 0.3 s). Models include transductive baselines, pre-trained inductive baselines, and fine-tuned variants (FT). Best performing method in **bold**, second best underlined. Results reported as mean $\pm$ standard deviation across 5 random seeds.

Model	MAE $\downarrow$	MSE $\downarrow$	MRE (%) $\downarrow$	PS F1 $\uparrow$
<i>Transductive</i>				
Mean	0.1851 $\pm$ 0.0006	0.0569 $\pm$ 0.0004	37.38 $\pm$ 0.11	0.000 $\pm$ 0.000
MF	<u>0.1321$\pm$0.0008</u>	<u>0.0318$\pm$0.0004</u>	<u>26.68$\pm$0.16</u>	<u>0.441$\pm$0.100</u>
<i>Inductive</i>				
RNN	0.1477 $\pm$ 0.0001	0.0329 $\pm$ 0.0000	29.83 $\pm$ 0.03	0.008 $\pm$ 0.007
Bi-RNN	0.1470 $\pm$ 0.0001	0.0326 $\pm$ 0.0000	29.69 $\pm$ 0.03	0.035 $\pm$ 0.018
TTS-Transformer	0.1473 $\pm$ 0.0001	0.0328 $\pm$ 0.0000	29.74 $\pm$ 0.03	0.074 $\pm$ 0.038
<i>Fine-tuned (FT)</i>				
RNN + FT	0.1469 $\pm$ 0.0001	0.0327 $\pm$ 0.0000	29.68 $\pm$ 0.03	0.000 $\pm$ 0.000
Bi-RNN + FT	0.1464 $\pm$ 0.0002	0.0325 $\pm$ 0.0000	29.57 $\pm$ 0.05	0.004 $\pm$ 0.004
TTS-Transformer + FT	0.1469 $\pm$ 0.0002	0.0330 $\pm$ 0.0001	29.68 $\pm$ 0.05	0.032 $\pm$ 0.014
FIBMAP	0.0706$\pm$0.0007	0.0104$\pm$0.0003	14.24$\pm$0.15	0.922$\pm$0.016

In contrast, the next best baseline method (MF) failed to capture the complex spatiotemporal AF dynamics, producing phase maps with substantial errors (Fig. 5.5A) and poor signal reconstruction (Fig. 5.5B).

Assessing Performance across Mapping Parameters and AF Organisation

Sensitivity analysis evaluated performance across catheter parameters and AF organisation levels (quantified by Shannon entropy; patients divided at median into low/high entropy groups). Fig. 5.6A shows reconstruction MAE across catheter areas (5–20% instantaneous surface coverage) and dwell times (0.5–4.0 s) with no spatial overlap. The low entropy group consistently achieved lower MAE (0.0728 average) compared to high entropy (0.0740 average). Both groups generally demonstrated improved performance with increased catheter area and reduced dwell time, achieving minimal MAE of 0.066 (low entropy, 20% area, 0.5 s dwell). However, this temporal sensitivity was less pronounced in the high entropy group, where reconstruction MAE became largely agnostic to dwell time variations once instantaneous coverage reached 15–20%. Shorter dwell times enabled faster atrial sampling, allowing regions of the atria to be visited multiple times by the catheter within the recording durations. Combinations where regions of the atria have not yet been visited within the recording duration (for example, because of low area and high dwell

Table 5.7: Phase singularity detection performance on human AF AcQMap test set across precision thresholds. Models evaluated using simulated sequential contact mapping (patches covering $\sim 10\%$ atrial surface area, 1 s dwell time, no spatial overlap). F1-scores reported at four spatiotemporal precision thresholds. Best performing method in **bold**, second best underlined. Results reported as mean $\pm$ standard deviation across 5 random seeds.

Method	High Precision (≈ 0.5 cm, 0.1 s)	Moderate Precision (≈ 1.0 cm, 0.2 s)	Standard Precision (≈ 1.5 cm, 0.3 s)	Lenient Precision (≈ 2.0 cm, 0.4 s)
<i>Transductive</i>				
Mean	0.000 $\pm$ 0.000	0.000 $\pm$ 0.000	0.000 $\pm$ 0.000	0.000 $\pm$ 0.000
MF	0.097 $\pm$ 0.021	<u>0.245$\pm$0.062</u>	<u>0.441$\pm$0.100</u>	<u>0.598$\pm$0.120</u>
<i>Inductive</i>				
RNN	0.006 $\pm$ 0.006	0.007 $\pm$ 0.007	0.008 $\pm$ 0.007	0.009 $\pm$ 0.008
Bi-RNN	0.020 $\pm$ 0.014	0.030 $\pm$ 0.016	0.035 $\pm$ 0.018	0.038 $\pm$ 0.020
TTS-Transformer	0.010 $\pm$ 0.008	0.039 $\pm$ 0.017	0.074 $\pm$ 0.038	0.149 $\pm$ 0.097
<i>Fine-tuned</i>				
RNN + FT	0.000 $\pm$ 0.000	0.000 $\pm$ 0.000	0.000 $\pm$ 0.000	0.000 $\pm$ 0.000
Bi-RNN + FT	0.004 $\pm$ 0.003	0.004 $\pm$ 0.003	0.004 $\pm$ 0.004	0.005 $\pm$ 0.004
TTS-Transformer + FT	0.004 $\pm$ 0.004	0.018 $\pm$ 0.008	0.032 $\pm$ 0.014	0.039 $\pm$ 0.016
FIBMAP	0.487$\pm$0.059	0.794$\pm$0.039	0.922$\pm$0.016	0.977$\pm$0.008

time) are not shown.

Performance was further assessed by imputation horizon, the spatiotemporal distance (in graph hops) from the nearest observed measurement. Fig. 5.6B demonstrates that MAE increased with imputation horizon, with the high entropy group showing a sharper rise before plateauing, possibly reflecting loss of phase coherence at longer distances. Concurrently, prediction interval width (10th–90th quantiles) increased with imputation horizon (Fig. 5.6B), with wider intervals indicating greater uncertainty at longer spatiotemporal distances. This uncertainty quantification is crucial for reliably interpreting whole-atrial imputation maps and making informed clinical assessments.

Interpreting FIBMAP’s State and Embedding Spaces

Fig. 5.7Ai shows hidden state trajectories for a representative node (visualised via t -SNE [Maaten and Hinton, 2008]), revealing orbital patterns and transitions characteristic of AF dynamics. The hidden states exhibit chaotic dynamics characteristic of AF, with trajectories diverging around fixed points. A recurrence plot [Marwan et al., 2007] of the state space trajectory (Fig. 5.7Aii) reveals structured dynamics, including parallel lines to the main diagonal at later times, a signature of deterministic systems where trajectories repeatedly progress through the same sequence of states. While comprehensive state space analysis exceeds this chapter’s scope, these findings demonstrate that FIBMAP’s architecture inherently captures important AF dynamical features that could enable patient electrophenotyping.

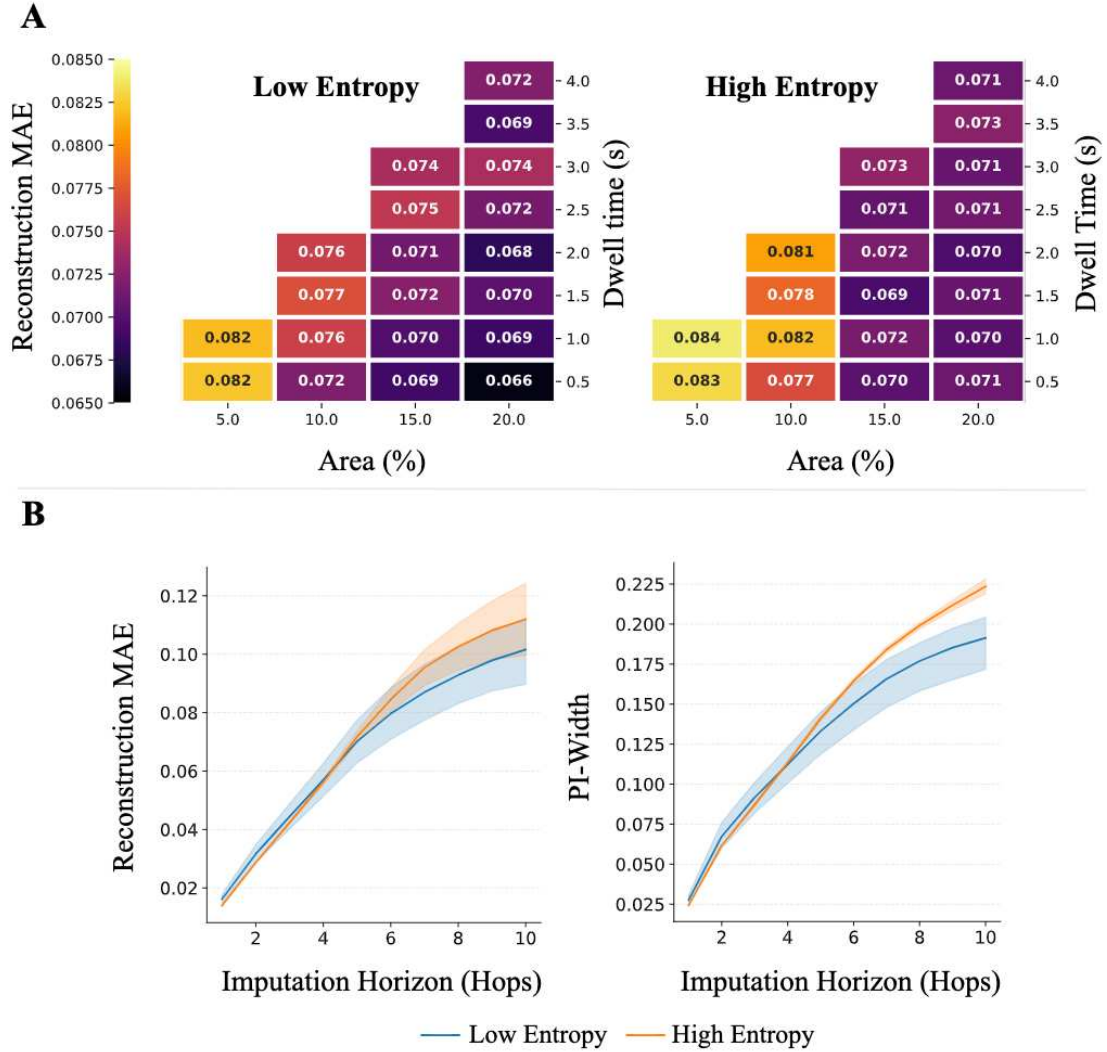


Figure 5.6: Sensitivity analysis of FIBMAP reconstruction performance. A) Mean absolute error (MAE) as a function of catheter surface area and dwell time, stratified by AF organisation (low vs. high entropy). Cells are only shown for combinations that sample measurements from all regions of the atria within the recording duration. B) MAE (left) and prediction interval width (PI-Width, right) versus imputation horizon (spatiotemporal distance in graph hops from nearest observation). Shaded regions indicate the standard deviation across test patients.

FIBMAP’s learned patient and node embedding parameters provide additional structural insights. Node embeddings visualised via t -SNE [Maaten and Hinton, 2008] (Fig. 5.7B) reveal patient-specific clusters while capturing dominant frequency, Shannon entropy, and voltage amplitude patterns. Notably, Fig. 5.7Biv reveals small clusters of low-voltage regions, indicating that the node embeddings may encode local voltage characteristics despite the min-max normalisation applied during preprocessing. This is particularly significant because in unobserved regions during clinical deployment, voltage normalisation cannot be reversed without knowledge of the true voltage range, yet FIBMAP’s embeddings indicate this substrate information may be captured implicitly. These embedding spaces could enable systematic comparison of tissue properties across patients and anatomical locations, while patient embeddings could identify individuals with similar global dynamics.

5.5.3 Translational Showcase: Contact Mapping with HD Grid

Having completed the main validation on optical mapping of 17 *ex vivo* rat hearts and non-contact mapping of 51 human AF patients, the extensibility of the approach to routine contact mapping is now assessed. In this supplemental experiment, the human AF FIBMAP model (pre-trained on non-contact recordings) is transferred to HD Grid recordings from three patients. Imputed phase signals were compared to non-contemporaneous AcQMap “ground truth” in sliding windows (0.5–4.0 s duration, 0.1 s stride), yielding 354,600–1,285,218 cross-correlation values per pair of patients at 0.5 s window duration, to assess if FIBMAP reconstructions captured similar patient-specific dynamical signatures as AcQMap.

Cross-correlation distributions for 0.5 s window duration revealed consistently higher intra-patient correlations (same patient; 99th percentiles: 0.16–0.17) versus inter-patient comparisons (different patients; 0.13–0.14) for patients i–ii (Fig. 5.8Bi–ii), demonstrating successful patient-specific pattern capture. In contrast, patient iii showed identical 99th percentiles (0.13) for both intra- and inter-patient comparisons (Fig. 5.8Biii). These patterns remained consistent across all sliding window durations (Fig. 5.8Ci–iii): intra-patient correlations consistently exceeded inter-patient correlations for patients i–ii, while for patient iii, inter-patient correlations were slightly higher. This lack of patient-specific discrimination in patient iii may reflect geometric registration uncertainty during this comparison, more disorganised dynamics, its shorter recording duration (13 vs. 18–20 minutes), or higher imputation error. Critically, all correlations (both intra- and inter-patient) substantially exceeded shuffled baselines (99th percentiles: ~ 0.01) by an order of magnitude across all three patients, confirming FIBMAP captured genuine AF dynamics rather than spurious correlations.

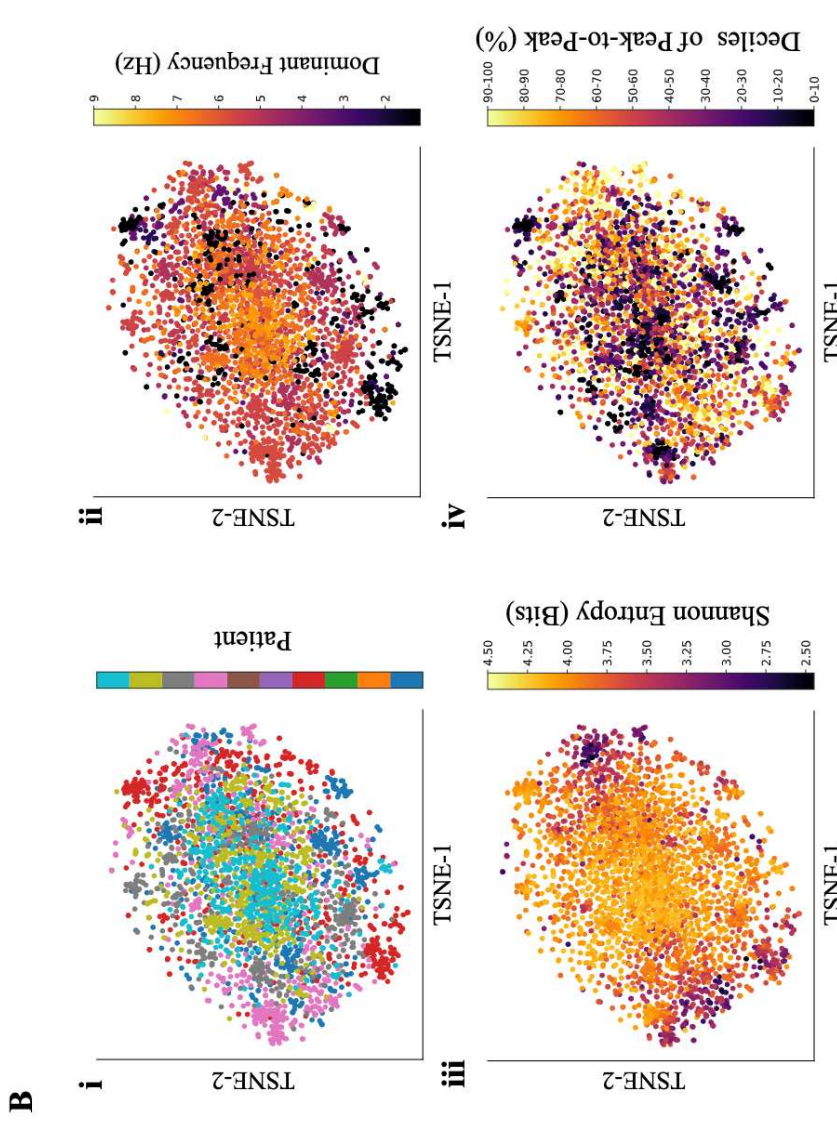
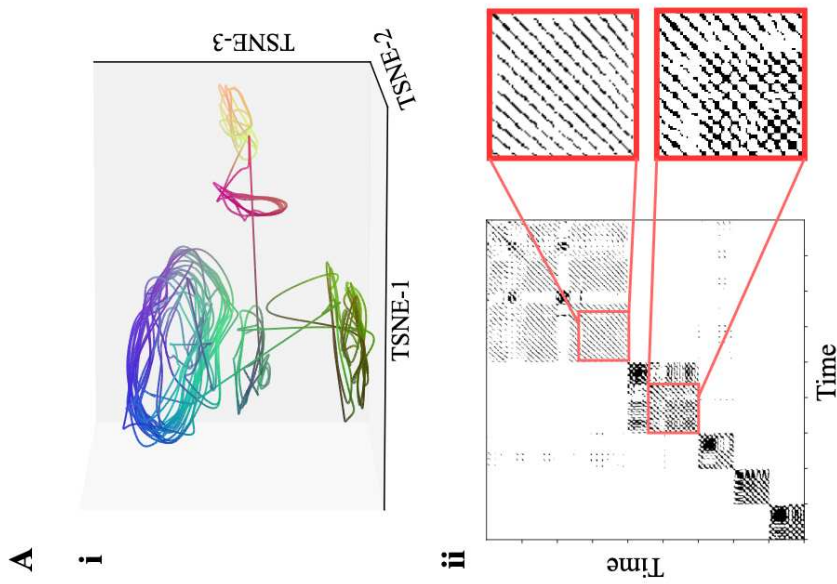


Figure 5.7: FIBMAP's learned state and embedding spaces. A) Hidden state dynamics for a representative node. (i) 3D state space trajectory visualised via t -SNE with RGB colour-coding revealing orbital patterns and transitions. (ii) Recurrence plot showing structured dynamics with parallel diagonal lines indicating deterministic progression through repeated state sequences. B) Node embeddings from all test patients visualised via t -SNE. (i) Embeddings coloured by patient identity, showing both patient-specific clustering and cross-patient relationships. (ii-iv) Node embeddings capture complex relationships with dominant frequency, Shannon entropy, and peak-to-peak voltage amplitude.

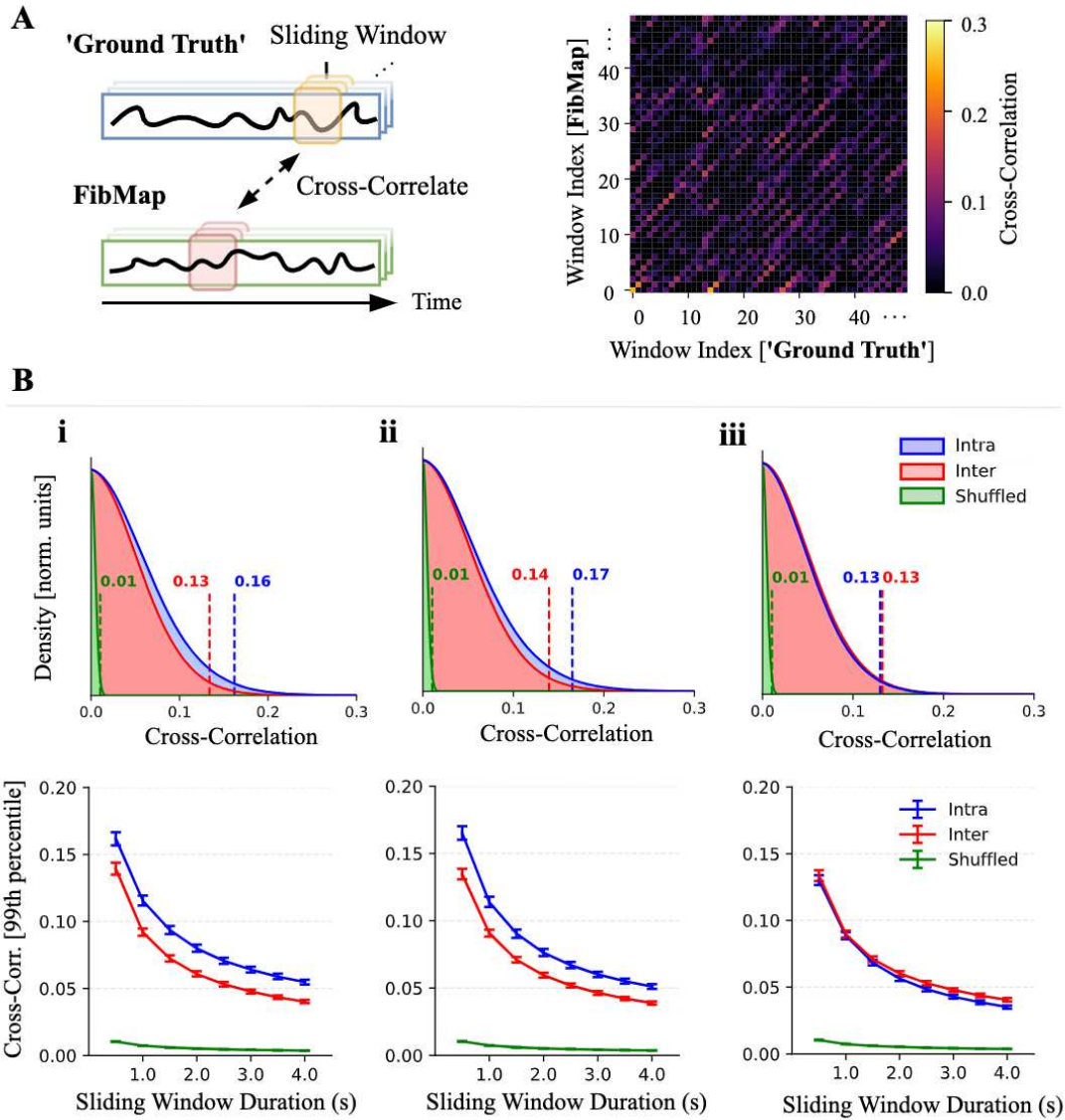


Figure 5.8: Proof-of-concept clinical translation of FIBMAP to real-world HD Grid contact mapping. A) Sliding window cross-correlation methodology (left) for comparing non-contemporaneous recordings by measuring phase signal similarity across temporal windows. Representative correlation matrix (right) shows pairwise comparisons across all window combinations. B) Cross-correlation distributions for three patients (i–iii) using 0.5 s sliding windows, comparing intra-patient (blue), inter-patient (red), and spatiotemporally shuffled (green) correlations. Top: kernel density estimates with 99th percentiles indicated. Patients i–ii demonstrate patient-specific pattern capture with higher intra-patient correlations (0.16–0.17) versus inter-patient (0.13–0.14). Patient iii shows overlapping distributions (both 0.13). All correlations exceed shuffled baseline (~ 0.01) by an order of magnitude, confirming reconstruction of genuine AF dynamics. Bottom: 99th percentile correlations versus sliding window duration (0.5–4.0 s), with error bars showing 95% confidence intervals from bootstrap sampling.

Fig. 5.9 shows representative results for each patient (i–iii): phase map snapshots from the highest-correlation 1-second window, alongside signal traces from three individual nodes showing the best, median, and worst local agreement within that window. Despite the non-time-aligned nature of recordings, FIBMAP reconstructions and AcQMap “ground truth” displayed similar patterns across all patients, including patient iii. The reconstructed phase maps reveal complex AF dynamics (multiple wavefronts, rotational activity) not visible in sparse HD Grid observations covering <10% of the atrial surface per time instant.

The findings of this supplemental study provide a showcase for the clinical translation of a pre-trained FIBMAP model, demonstrating its capacity to reconstruct personalised, whole-atrial dynamics from routine contact mapping data. However, several sources of uncertainty remain, including the comparison of non-contemporaneous recordings, geometric projection requirements between recording modalities, and a small clinical sample size. Addressing these factors is essential for the further refinement of this translational approach and for establishing the requirements for future large-scale prospective validation.

5.6 Discussion

The inability to map AF effectively during invasive procedures limits the application of tailored ablation strategies and potentially hinders curative treatment outcomes. This chapter introduced imputation mapping to reconstruct complete global AF dynamics from recordings collected through localised sequential contact mapping, which observes only a fraction of the atrium at any time. While cumulative coverage grows over time as regions are sampled during the clinical mapping procedure, the value of FIBMAP lies in providing a simultaneous global view from those sparse sequential measurements. Until now, these sparse measurements could not be effectively integrated into global maps due to the disorganised nature of AF (Sec. 3.3). By modelling cardiac activation as a non-linear state space using a GRNN (Sec. 2.3.2), the results demonstrate that accurate, personalised reconstruction of whole-atrial dynamics is achievable from sparse measurements through graph-based spatiotemporal machine learning.

Core validation. The imputation mapping proof-of-concept was rigorously conducted on two complementary datasets: high-fidelity optical mapping of rat VF (17 hearts, over 253 million spatiotemporal data points) and clinical non-contact mapping from 51 persistent AF patients (over 4.0 billion spatiotemporal data points). The results demonstrated that FIBMAP substantially outperformed baseline imputation methods, achieving 53% improvement in MAE and 23% improvement in PS detection on VF optical mapping, and 47% improvement in MAE and 109% improvement

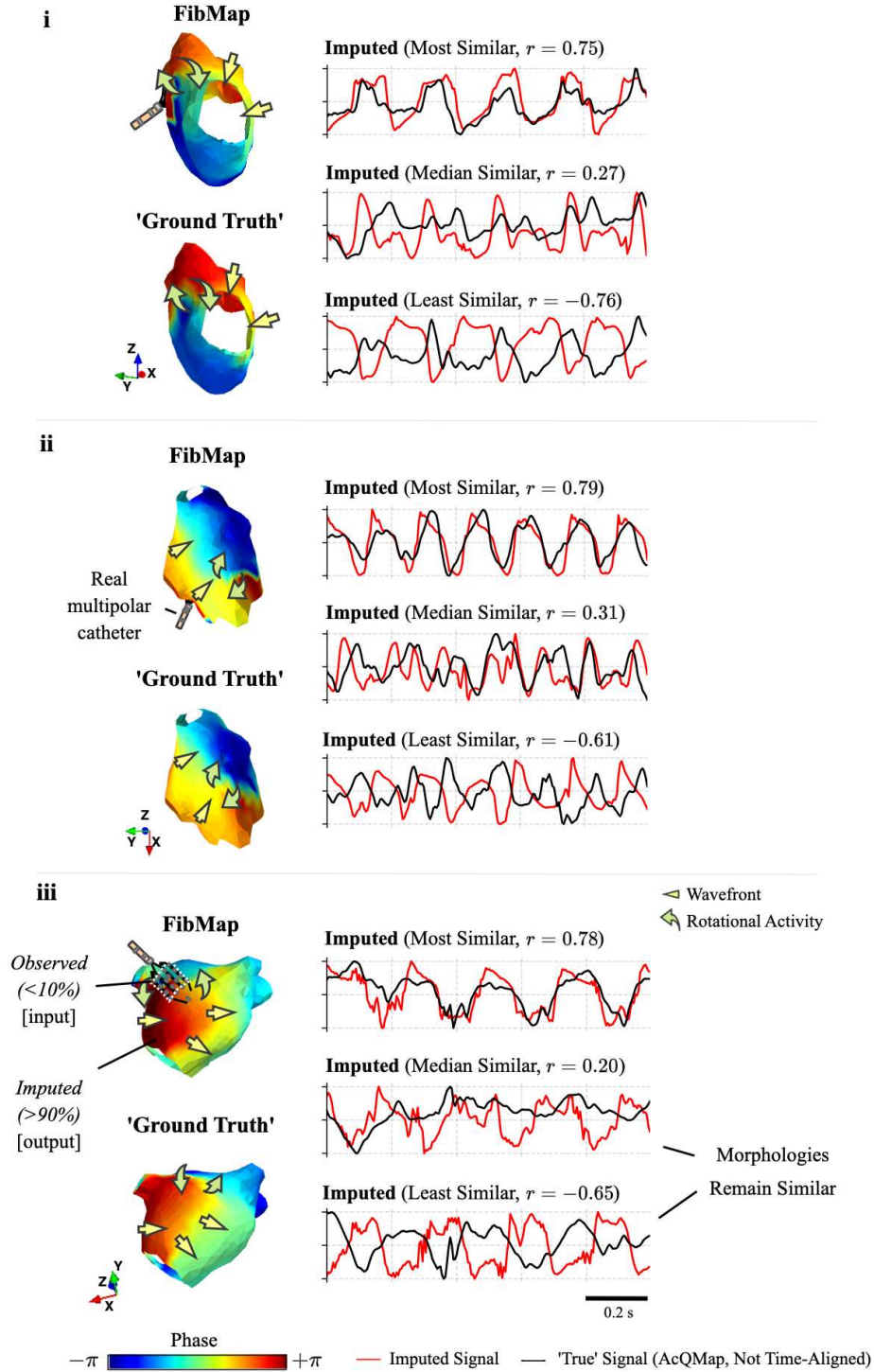


Figure 5.9: Representative FIBMAP reconstructions from HD Grid contact mapping for three patients (i–iii). Left: Phase map snapshots from the highest-correlation 1-second window, comparing FIBMAP reconstructions (top) with AcQMap “ground truth” (bottom). FIBMAP captures complex AF dynamics including multiple wavefronts (yellow arrows) and rotational activity (green arrows) from sparse observations instantaneously covering <10% atrial surface. Right: Signal traces from three individual nodes showing best (top), median (middle), and worst (bottom) local agreement within the displayed window, with correlation coefficients indicated. Despite non-contemporaneous recordings, FIBMAP reconstructions demonstrate similar morphologies and dynamical patterns to AcQMap “ground truth” across all patients, though phase jittering occurs in some regions (patient iii, bottom trace).

in PS detection on human AF data (observing only 10% of the ventricular and atrial surfaces, respectively). FIBMAP accurately captured both amplitude and phase relationships across space and time, and reconstructed complex fibrillation features (rotational activity, multiple wavefronts, and phase singularities) that were invisible in the sparse observations. Uncertainty was quantified through prediction intervals that reflected reconstruction confidence, where intervals widened in regions where observations were temporally or spatially distant.

The substantial performance improvement of FIBMAP over the baselines warrants discussion. The purely temporal baselines (RNN, Bi-RNN) process each node’s time series independently, exploiting temporal structure but ignoring the spatial relationships that govern wavefront propagation. The TTS-Transformer introduces spatial processing but factorises temporal and spatial dimensions, limiting the fine-grained spatiotemporal coupling that T&S models achieve ([Sec. 2.3.2](#)). MF, while capturing global spatial patterns through low-rank approximation, assumes linearity, which is violated by the non-linear dynamics of fibrillation. FIBMAP’s GRNN architecture, by interleaving graph message passing with recurrent temporal processing at every time step, provides the fine-grained spatiotemporal coupling necessary to model the non-linear wave propagation dynamics of fibrillation. The addition of patient-specific and node-specific embeddings further differentiates FIBMAP from baselines by enabling personalised reconstruction through the hybrid global-local strategy discussed in [Sec. 2.3.3](#).

Embedding spaces and electrophenotyping. Visualisation of FIBMAP’s hidden state and embedding spaces revealed structure with potential implications for AF electrophenotyping. Hidden states exhibited chaotic dynamics characteristic of fibrillation and recurrence plots revealed deterministic structure. Node embeddings captured complex, non-linear relationships with dominant frequency, Shannon entropy, and peak-to-peak voltage amplitude, with visible clustering of low-voltage regions despite min-max normalisation applied during preprocessing. This suggests embeddings may encode substrate information, which is particularly relevant since reversing voltage normalisation in unobserved regions requires knowledge of the true voltage range. The normalisation approach adopted here was designed to emphasise propagation patterns rather than absolute voltage amplitudes; however, raw sparse contact measurements remain available for standard substrate mapping, allowing FIBMAP’s propagation maps to be interpreted alongside established voltage-based approaches [[Rivera et al., 2024](#)]. Comprehensive interpretation and clinical validation of embedding information requires future studies with follow-up data, tissue characterisation, and outcome measures to assess potential for predicting ablation response or guiding treatment strategies.

Sensitivity to mapping parameters. The sensitivity analysis demonstrated that reconstruction performance varied systematically with catheter parameters and AF organisation, achieving optimal results (MAE 0.066) with 20% instantaneous coverage, 0.5 s dwell times, and more organised dynamics. These relationships could inform future adaptive catheter guidance algorithms that tailor mapping strategies to individual patient characteristics and real-time AF organisation assessment.

Showcase of clinical translation. Beyond the core validation presented in this study, the supplemental showcase demonstrated the clinical translation of FIBMAP to routine clinical recordings with the HD Grid mapping catheter through the pre-training and fine-tuning procedures developed here. The results demonstrated clinical translation of FIBMAP to routine recordings, with the cross-correlation analysis between FIBMAP reconstructions and AcQMap ground truth revealing patient-specific pattern capture in two of three patients and all patients showing correlations exceeding chance baselines by an order of magnitude. However, one patient showed no distributional discrimination between intra- and inter-patient correlations, highlighting the challenges of conducting a robust validation using non-contemporaneous recordings, where AF’s inherent non-stationarity may compound assessment difficulties. Future large-scale prospective studies with contemporaneous contact mapping recordings and a suitable global ground truth will enable direct temporal validation and establish the method’s robustness across diverse patient phenotypes and AF organisation levels.

Relationship to existing mapping and modelling approaches. Unlike digital twin approaches using pre-procedural MRI [Boyle et al., 2019; Sakata et al., 2024], FIBMAP reconstructs actual intra-procedural AF dynamics without requiring specialised imaging, enabling real-time visualisation complementary to prospective planning tools. Additionally, while whole-atrial mapping methods exist, such as AcQMap and non-invasive electrocardiographic imaging, these non-contact approaches have limitations in spatial resolution and clinical adoption due to cost, procedural complexity, and the ill-posed nature of inverse reconstruction (Sec. 3.5) [Narayan, Krummen, et al., 2012; Buch et al., 2016; Rudy and Burnes, 1999; Roney et al., 2017]. FIBMAP generates whole-atrial maps from the sparse recordings collected routinely, without the hardware, cost, or inverse-problem constraints of these existing modalities. A natural question for future work is whether models trained on diverse data sources can eventually *exceed* the accuracy ceiling of any single training modality.

The translational showcase demonstrated that pre-trained models generalise across recording modalities via fine-tuning, suggesting that FIBMAP is not tightly bound to any single training source.

This opens a promising direction: a model trained on a diverse synthesis of high-resolution optical mapping, large-scale clinical non-contact data, and biophysical simulations could yield a stronger predictive framework than any single modality alone [Y. Huang et al., 2021]. By triangulating between data sources, such a model could learn to filter modality-specific noise and artefacts, potentially surpassing the accuracy of the individual recording tools used for training. The optimal composition of training data and whether such multi-modal strategies can ultimately improve clinical outcomes beyond current mapping approaches are important questions for prospective validation.

Broader implications. Imputation mapping provides comprehensive spatiotemporal data previously inaccessible during clinical procedures, offering clinicians a simultaneous global view of AF dynamics from the same sparse contact recordings already collected routinely. This global context enables visual identification of driver regions, supports electrophenotyping based on each patient’s unique AF dynamics [F. S. Ng et al., 2020], and may enhance automated annotation algorithms that currently operate on local electrogram features alone [Liao et al., 2021; Seitz et al., 2022; Manongi, J. Kim, and Goldbarg, 2023]. Prospective validation will determine whether the additional spatial information translates into improved identification of ablation targets that conventional sequential mapping may miss.

5.7 Summary and Relation to Thesis

This chapter demonstrated that relational inductive biases, introduced through graph-based representations of the cardiac surface, can enable the reconstruction of global fibrillation dynamics from sparse sequential measurements. FIBMAP achieves this by combining a GRNN architecture that interleaves graph message passing with recurrent temporal processing, a bidirectional encoder that exploits both past and future observations, and a hybrid global-local parameterisation that personalises reconstruction through patient-specific and node-specific embeddings. The approach was validated on two complementary datasets spanning different species, arrhythmia types, and recording modalities, with a supplemental translational showcase on routine clinical contact mapping.

This chapter directly addresses the first research question of this thesis (**RQ1**): *can graph-based spatiotemporal machine learning reconstruct global cardiac fibrillation dynamics from sparse clinical measurements?* The results demonstrate that the graph-based spatiotemporal framework developed in Chapter 2 can be instantiated as a practical tool for imputation mapping. The GRNN architecture of FIBMAP is a concrete instance of the T&S STGNN paradigm presented

in Sec. 2.3.2, with the anisotropic message-passing mechanism (Eq. 5.3.2) instantiating the general MPNN framework of Fig. 2.2.5, the bidirectional encoder addressing the imputation-specific design considerations discussed in Sec. 2.3.2, and the hybrid global-local embedding strategy following the approach motivated in Sec. 2.3.3.

Two open challenges identified in Sec. 2.3.3 remain. First, while FIBMAP’s quantile regression provides prediction intervals, these are model-dependent estimates that lack formal coverage guarantees. The quantile outputs reflect the model’s learned conditional distribution but do not account for distribution shift between training and deployment conditions, nor do they provide the calibrated, distribution-free coverage guarantees desirable in clinical settings. Chapter 6 addresses this gap by developing a conformal prediction framework that provides post-hoc, distribution-free prediction intervals with finite-sample coverage guarantees for spatiotemporal graph predictions, exploiting relational structure in prediction residuals to produce tighter intervals than methods that treat each time series independently.

Second, FIBMAP operates on a static anatomical graph, $\mathcal{G}$, derived from each patient’s chamber surface geometry. While the anisotropic edge weights, $\alpha_{ij}^{(l)}$, learned within each message-passing layer weight incoming messages from neighbouring nodes in a data-adaptive manner (as a function of both source and target node representations), they have the capacity to implicitly capture aspects of functional connectivity. However, these weights are learned independently for each layer and each gate within the GRNN, making their interpretation as a coherent picture of the underlying conduction pathways challenging. A more interpretable alternative is to learn the graph topology, $\mathcal{G}$, directly from the time series observed, or now reconstructed, at each node, yielding an explicit and parsimonious representation of functional connectivity. Chapter 7 pursues this direction by developing an online algorithm for time-varying graph topology learning that extends the causal graph process framework of Sec. 2.3.1 to the adaptive setting, enabling the tracking of evolving functional connectivity from streaming spatiotemporal data. Together, the three contributions of this thesis are designed to form a progressive pipeline: FIBMAP *reconstructs* global dynamics from sparse measurements, CoRel (Chapter 6) *analyses* the reliability of these predictions through calibrated uncertainty quantification, and AdaCGP (Chapter 7) *characterises* the fibrillation mechanism by learning the dynamic functional connectivity that organises the reconstructed dynamics.

It is worth noting that the diffusion-based generative models demonstrated in Chapter 4 for temporal imputation may also offer advantages in the spatiotemporal reconstruction setting. Recent work has increasingly explored the integration DDPMs with graph-based spatiotemporal architectures for imputation tasks [M. Liu et al., 2023; Hu et al., 2024; Yakun Chen et al., 2024; Z. Liu,

X. Zhao, and You Song, [2025](#)], combining the distributional flexibility of diffusion models with the relational inductive biases of GNNs. Extending FIBMAP to operate as a conditional diffusion model represents a promising direction for future work, with the potential to improve both reconstruction fidelity and uncertainty quantification.

Part III

Analysis and Characterisation of Spatiotemporal Systems

Chapter 6

Relational Conformal Prediction for Calibrated Uncertainty Quantification

This chapter addresses the problem of uncertainty quantification for spatiotemporal forecasting by exploiting observations at correlated time series through a graph-based conformal prediction framework. The preceding chapter demonstrated that graph-based spatiotemporal models can reconstruct global fibrillation dynamics from sparse measurements, but the resulting predictions lack formal coverage guarantees. Here, COREL (Conformal Relational Prediction) is introduced: a distribution-free, post-hoc method that constructs prediction intervals for any pre-trained forecasting model by training a spatiotemporal graph neural network on calibration residuals to approximate the quantile function of the error distribution. The relational structure among time series is assumed sparse and is learned end-to-end from the observed residuals, allowing the method to condition uncertainty estimates on observations at neighbouring nodes without requiring the graph to be known a priori. An adaptive component handles non-stationary error distributions by updating local node embeddings over time. Validated on traffic, air quality, and energy consumption benchmarks, COREL achieves state-of-the-art uncertainty quantification with accurate coverage guarantees while maintaining computational scalability. While the method is validated in the forecasting setting, extending it to the imputation setting to provide calibrated prediction intervals for reconstruction methods such as FIBMAP is an identified direction for future work, discussed in [Sec. 6.5](#).

6.1	Introduction	133
6.2	Preliminaries	135
6.3	Methodology	138
6.4	Experiments	145
6.5	Discussion	153
6.6	Summary and Relation to Thesis	154

This chapter is based on Cini, Jenkins, et al. (2025).

6.1 Introduction

The preceding chapter introduced FIBMAP, a graph recurrent neural network that reconstructs global fibrillation dynamics from sparse sequential catheter measurements (Chapter 5). While FIBMAP’s quantile regression outputs provide prediction intervals, these estimates are model-dependent and lack formal coverage guarantees; they reflect the model’s learned conditional distribution but offer no assurance that the stated confidence level is achieved in practice. If imputation maps are to guide clinical decision-making in the future, a more rigorous treatment of uncertainty is essential. Identifying regions of high predictive confidence could support more targeted ablation strategies, while flagging unreliable reconstructions is equally important to avoid misguided treatment decisions. More broadly, the challenge of providing calibrated uncertainty for spatiotemporal predictions extends well beyond cardiac mapping. Any application where forecasts inform consequential decisions, from energy load management to traffic routing, requires not only accurate point estimates but also reliable measures of their uncertainty.

Uncertainty quantification methods [R. C. Smith, 2024; Vovk, Gammerman, and Shafer, 2005] address this need by providing prediction intervals on the forecasting error, enabling practitioners to assess the reliability of individual predictions rather than relying on aggregate accuracy metrics [Makridakis, 1996; Petropoulos et al., 2022]. Crucially, the inter-series dependencies that spatiotemporal graph neural networks exploit for prediction (Chapter 2) also structure the prediction errors: overlooking these dependencies when quantifying uncertainty can lead to unnecessarily wide prediction intervals, while exploiting them may provide additional information to tighten the intervals [Zambon and Alippi, 2022].

As discussed in Sec. 2.3.3, existing probabilistic forecasting frameworks often rely on strong distributional assumptions and require substantial modifications to the base predictor, precluding their application to pre-trained models. Among deep learning approaches, generative architectures [Salinas et al., 2020; Rangapuram et al., 2018; Bézenac et al., 2020; Rasul et al., 2021] and methods based on quantile regression [R. Wen et al., 2017; Gasthaus et al., 2019; Kan et al., 2022; Gouttes et al., 2021] have been proposed for probabilistic forecasting [Hyndman and Athanasopoulos, 2018; Petropoulos et al., 2022; Benidis et al., 2022], but remain tightly coupled to a specific model class. Within the spatiotemporal graph neural network literature specifically, the options are equally constrained. D. Wu et al. (2021) combine spatiotemporal graph neural networks with standard deep learning uncertainty quantification techniques, Pal et al. (2021) embed a graph-based state-space model within a Bayesian framework, H. Wen et al. (2023) couple spatiotemporal graph neural networks with denoising diffusion [J. Ho, Jain, and Abbeel, 2020],

and Zambon, Cini, et al. (2023) design probabilistic graph state-space models [M. Jin et al., 2024; Cini, Marisca, Zambon, et al., 2025]. All require training the uncertainty mechanism jointly with the predictor and cannot be applied post-hoc to an existing model.

Conformal prediction [Vovk, Gammerman, and Shafer, 2005; Angelopoulos, Bates, et al., 2023] offers a compelling alternative: a distribution-free framework that constructs prediction intervals with marginal coverage guarantees from observed prediction residuals, applicable post-hoc to any pre-trained forecasting model. Standard conformal prediction requires that calibration and test data are exchangeable, yet this condition is violated in the temporal setting where successive observations are correlated. Recent work has made considerable progress in extending the framework beyond this requirement [Stankeviciute, M Alaa, and Schaar, 2021; C. Xu and Y. Xie, 2023a; C. Xu and Y. Xie, 2023b; Gibbs and Candes, 2021; R. F. Barber et al., 2023; Auer et al., 2023], as discussed in Sec. 6.2.2. However, all existing conformal prediction methods for time series operate independently on each target time series and do not account for relational dependencies among the series when constructing prediction intervals.

This chapter fills this gap by introducing COREL, a conformal prediction method that leverages graph deep learning operators (Fig. 2.2.5) to exploit relational structure in prediction residuals. A spatiotemporal graph neural network (Sec. 2.3.2) is trained on a calibration set to approximate the quantile function of the error distribution, conditioning uncertainty estimates at each node on recent observations at neighbouring nodes as defined by a learned sparse graph topology. The method does not require the relational structure to be known *a priori* and includes an adaptive component to handle non-stationary error distributions. The main contributions of this chapter are threefold: first, the introduction of graph deep learning to conformal prediction for time series; second, a principled and effective method for quantifying uncertainty from observations across a collection of correlated time series; and third, a family of graph-based architectures for uncertainty estimation that share most learnable parameters among time series while including node-level parameters that dynamically adapt to changes in each target sequence.

It is emphasised that this chapter operates in the *forecasting* setting, where the goal is to quantify the uncertainty of predictions about future values. While potential extensions to the imputation setting are discussed in Sec. 6.5, the method and experiments presented here focus on the standard multi-step-ahead forecasting problem.

6.2 Preliminaries

This section introduces the problem setting and the foundational concepts on which COREL builds, drawing on the conformal prediction and quantile regression frameworks outlined in [Sec. 2.3.3](#) and the spatiotemporal graph neural network architectures developed in [Sec. 2.3.2](#).

6.2.1 Problem Formulation

Consider a collection of N correlated time series. The scalar target variable associated with the i -th time series at time step, t , is denoted $x_i[t] \in \mathbb{R}$. The vector $\mathbf{X}[t] \in \mathbb{R}^{N \times 1}$ stacks the target variables across the entire collection, $\mathbf{X}[t:t+T]$ represents the sequence within the interval $[t, t+T)$, and $\mathbf{X}[<t]$ holds observations up to (but excluding) time step t . The time series are assumed to be homogeneous, meaning that all variables describe the same physical quantity (e.g. voltage or temperature). Exogenous covariates associated with each time series are denoted, $\mathbf{U}[t] \in \mathbb{R}^{N \times d_u}$, where d_u is the covariate dimension.

It is assumed that the i -th time series is generated by a stochastic, time-invariant process,

$$x_i[t] \sim p(x_i[t] \mid \mathbf{X}[<t], \mathbf{U}[<t]), \quad (6.1)$$

and hypothesise the existence of a sparse predictive causality of the Granger type [Granger, 1969]. That is, the values of each time series depend on only a small subset of other series in the collection. This sparsity assumption connects to the causal graph process framework developed in [Sec. 2.3.1](#), where the graph shift operator encodes the directed causal relationships among multivariate time series. The extension to collections of multivariate time series is orthogonal to the approach presented here [Feldman, Bates, and Romano, 2023]; the univariate case is considered to maintain a contained scope.

Forecasting. A trained parametric model is considered that produces point forecasts by predicting the H -steps-ahead observation, $\mathbf{X}[t+H]$, where $(H \geq 1)$, given a window of $W \geq 1$ past observations, $\mathbf{X}[t-W:t]$, and associated exogenous variables, $\mathbf{U}[t-W:t]$, such that

$$\hat{\mathbf{X}}[t+H] = f_{\boldsymbol{\theta}}(\mathbf{X}[t-W:t], \mathbf{U}[t-W:t]), \quad (6.2)$$

where $f_{\boldsymbol{\theta}}$ denotes a generic parametric forecasting model (e.g. a recurrent neural network for univariate time series, or a spatiotemporal graph neural network that additionally exploits inter-series dependencies). The formulation extends naturally to multi-step predictions, $\hat{\mathbf{X}}[t:t+H]$, but

single-step forecasting at horizon, H , is considered to simplify the presentation.

Uncertainty quantification. Given the trained point predictor, the objective is to construct prediction intervals, $C_{i,t}^\alpha$, such that

$$P(x_i[t+H] \in C_{i,t}^\alpha) \geq 1 - \alpha, \quad (6.3)$$

where α is the desired miscoverage level. When (6.3) holds, the prediction interval is said to achieve marginal coverage at level $1 - \alpha$. A stronger guarantee is conditional coverage,

$$P(x_i[t+H] \in C_{i,t}^\alpha \mid \mathbf{X}[<t], \mathbf{U}[<t]) \geq 1 - \alpha, \quad (6.4)$$

which conditions on the observed history and is generally harder to achieve [Angelopoulos, R. F. Barber, and Bates, 2024]. Among uncertainty quantification methods, the focus is on post-hoc approaches that construct prediction intervals for any given pre-trained point predictor, f_θ , without modifying the base forecasting architecture.

6.2.2 Conformal Prediction for Time Series

Standard conformal prediction methods [Vovk, Gammerman, and Shafer, 2005; Angelopoulos, Bates, et al., 2023] are distribution-free uncertainty quantification techniques that construct prediction intervals from empirical quantiles of conformal scores, scalar measures of the discrepancy between a prediction and the corresponding true observation. In the forecasting setting, the prediction residuals serve as conformal scores,

$$r_i[t] = x_i[t] - \hat{x}_i[t], \quad (6.5)$$

and write $\mathbf{R}[t] \in \mathbb{R}^{N \times 1}$ for the residual vector across the collection.

Split conformal prediction. Split conformal prediction (SCP) [Vovk, Gammerman, and Shafer, 2005] partitions the available data into a training set, used to fit the point predictor, f_θ , and a calibration set on which conformal scores are computed, making it a post-hoc approach. Given prediction residuals in the calibration set, $\mathcal{R}^{\text{cal}} = \{r_i[t]\}$, prediction intervals are constructed by taking empirical quantiles of the score distribution. For a desired miscoverage level α , the interval is centred on the point prediction and its width is determined by the $\lceil(1 - \alpha)\rceil$ -th quantile of the absolute calibration residuals. Under the assumption that calibration and test scores are drawn exchangeably from the same distribution, this procedure achieves the marginal coverage guarantee

of (6.3) [Angelopoulos, R. F. Barber, and Bates, 2024].

From marginal to conditional coverage. Marginal coverage provides only an average-case guarantee. SCP applies a single, globally calibrated interval width uniformly across all time steps and all nodes, yet in practice the conformal scores may exhibit both temporal and spatial structure. The error distribution across the collection may exhibit spatial and temporal dependencies, with certain times and series being inherently more volatile than others. Under such dependencies and heterogeneity, marginal coverage permits intervals that are excessively wide where uncertainty is low and insufficiently wide where it is high, so long as the coverage rate averages to the nominal level across the collection. The stronger notion of conditional coverage (6.4) requires the guarantee to hold within each specific context and is therefore the more operationally relevant target. Conditional coverage is, however, generally unattainable without explicitly modelling the dependencies in the score distribution [Angelopoulos, R. F. Barber, and Bates, 2024], requiring methods that learn this structure and adapt interval widths accordingly.

Extensions beyond exchangeability. Most standard conformal prediction methods assume that calibration and test data are exchangeable, meaning their joint probability distribution is invariant to permutation [Angelopoulos, R. F. Barber, and Bates, 2024]. This assumption is violated for time series data, where temporal ordering is fundamental [R. F. Barber et al., 2023]. Several recent extensions address this limitation. R. F. Barber et al. (2023) showed that approximate coverage can be achieved by reweighting residuals to account for non-exchangeability between calibration and test sets. Auer et al. (2023) learn such a reweighting scheme through an attention-based architecture. C. Xu and Y. Xie (2023b) introduce a sequential conformal prediction method that fits a quantile random forest [Meinshausen and Ridgeway, 2006] on recent prediction residuals at each time step. While these methods capture temporal dependencies in the residual distribution and move toward conditional coverage by adapting intervals over time, they operate on each time series independently and cannot exploit relational structure across correlated series.

A limited number of works have considered spatial structure within conformal prediction. Mao, R. Martin, and Reich (2024) develop a conformal framework for spatially correlated data in a static, non-temporal setting, while H. Jiang, Y. Xie, and Qiu (2024) quantify uncertainty in power outage prediction by fitting a quantile random forest on time series from neighbouring geographical units. However, neither approach addresses the full spatiotemporal forecasting problem on arbitrary relational structures.

No previous conformal prediction method has been designed to specifically operate on collections

of correlated time series using graph deep learning operators. Conformal prediction methods exist for multivariate time series [C. Xu, H. Jiang, and Y. Xie, 2024; S. H. Sun and R. Yu, 2024; Feldman, Bates, and Romano, 2023], but operate on a single multidimensional time series rather than exploiting inter-series relational structure. Although global-local models are common in forecasting architectures [Smyl, 2020; Benidis et al., 2022], this design choice has not been leveraged for spatiotemporal conformal prediction. Additionally, conformal prediction methods have been applied to static graphs for quantifying uncertainty of graph neural networks in inductive [Zargarbashi, Antonelli, and Bojchevski, 2023] and transductive [K. Huang et al., 2024] settings, often assuming node or edge exchangeability. These methods are limited to node classification [Clarkson, 2023] or link prediction [T. Zhao, Kang, and L. Cheng, 2024] and do not address the spatiotemporal forecasting setting.

6.3 Methodology

The objective is to construct prediction intervals by exploiting relational dependencies across the residuals of correlated time series. These dependencies are modelled as edges of a graph and learned under the assumption that the relational structure is sparse, which both reduces computational cost and acts as an inductive bias on the structure learning module. A spatiotemporal graph neural network is trained on the residuals of the calibration set to predict the quantiles of the error distribution [Koenker and Hallock, 2001]. By conditioning the prediction on the recent history at related time series, the method accounts for the dependency structure when estimating uncertainty, which is a key consideration when applying conformal prediction to non-exchangeable data [R. F. Barber et al., 2023]. Compared to existing methods [C. Xu and Y. Xie, 2023b; Auer et al., 2023] that only capture temporal dependencies, COREL additionally models spatiotemporal dependencies among different time series. COREL can be positioned among conformal prediction methods that learn a model of the conformal score distribution [C. Xu and Y. Xie, 2023b; Junghwan Lee, C. Xu, and Y. Xie, 2024], but differs from existing approaches by conditioning estimates on errors at both the target series and neighbouring nodes.

The remainder of this section is organised as follows. Subsection 6.3.1 presents the relational quantile predictor that forms the core of the approach. Subsection 6.3.2 describes how the graph structure is learned directly from data. Subsection 6.3.4 introduces the adaptive component for handling non-stationarity. Finally, Sec. 6.3.3 discusses theoretical properties and computational complexity.

6.3.1 Relational Quantile Predictor

Consider a standard SCP setup, where the available data are partitioned into training and calibration sets. The training set is used to fit the point predictor, f_{θ} , while the prediction residuals on the calibration set, $\mathcal{R}^{\text{cal}}$, are used to learn the quantile function of the error distribution. For the moment, spatial dependencies are encoded in a fixed adjacency matrix, $\mathbf{A} \in \mathbb{R}^{N \times N}$; the procedure for learning this matrix from data (if it is unknown for the task at hand) is developed in [Sec. 6.3.2](#).

Relational quantile regression. The core of COREL is a quantile regressor [Koenker and Hallock, 2001] that estimates the conditional quantile function of the residual distribution, exploiting relational structure among time series through a spatiotemporal graph neural network. Quantile regression estimates the quantile function from observations by minimising the pinball loss, which was introduced in [Sec. 5.3.3](#) as the training objective for FIBMAP’s imputation outputs. Here, the same principle is applied not to raw observations but to prediction residuals. In this way, the quantile regressor learns the conditional distribution of forecasting errors, enabling the construction of prediction intervals whose width adapts to the local error structure.

The quantile regressor is implemented as a hybrid global-local spatiotemporal graph neural network, following the design rationale discussed in [Sec. 2.3.3](#). This design shares most learnable parameters across all time series, reducing sample complexity, while a learnable node embedding, $\mathbf{v}_i \in \mathbb{R}^{d_q}$, associated with each time series provides local, target-specific adaptation [Smyl, 2020; Cini, Marisca, Zambon, et al., 2024]. Concretely, the architecture follows the TTS template of [Sec. 2.3.2](#), utilising a single-layer GRU as the temporal encoder, whose output is then fed into two message-passing layers analogous to those of Satorras, Rangapuram, and Januschowski (2022). The readout is a MLP that maps the learned representations to predictions of 39 equally spaced quantiles, approximating the full quantile function. For the CORNN ablation used in the experiments, the same architecture is employed but with the message-passing layers and node embeddings, $\mathbf{V} \in \mathbb{R}^{N \times d_q}$, removed, and the GRU permitted to have more than one layer. In addition to the residuals, date and time variables (when available) are encoded for each time step, which alongside the target time series values over the input window, are provided as covariates. The model is composed of the following processing layers

$$\mathbf{h}_i^{(0)}[t] = \text{Enc}(r_i[t-1], \mathbf{v}_i), \quad (6.6)$$

$$\mathbf{Z}[t] = \text{STGNN}\left(\mathbf{H}^{(0)}[\leq t], \mathbf{A}\right), \quad (6.7)$$

$$\hat{q}_i^{\alpha}[t+H] = \text{QDec}(\alpha, \mathbf{z}_i[t], \mathbf{v}_i), \quad (6.8)$$

where $r_i[t-1]$ is the prediction residual (6.5) and $\hat{q}_i^\alpha[t+H]$ is the predicted α -quantile at time step $t+H$ for the i -th time series. The encoder, $\text{Enc}(\cdot)$, is any encoding layer (e.g. a linear transformation or MLP). The $\text{STGNN}(\cdot)$ block instantiates the TTS template of (2.59)–(2.60), with the GRU as the sequence encoder and the message-passing layers operating on the graph $\mathbf{A}$ (Fig. 2.2.5). The quantile decoder, $\text{QDec}(\cdot)$, takes the quantile level α as an explicit input, following the implicit quantile network design [Dabney et al., 2018; Gouttes et al., 2021], and maps the spatiotemporal representation at each node to the predicted quantile at the specified level.

The family of quantile networks defined in (6.6)–(6.8) are referred to as relational quantile networks and written as

$$\hat{\mathbf{Q}}^\alpha[t] = \mathcal{Q}_\Psi(\alpha, \mathbf{V}; \mathbf{R}[t-W:t], \mathbf{A}), \quad (6.9)$$

where $\mathcal{Q}_\Psi$ denotes the shared (global) part of the network with parameters Ψ , and $\hat{\mathbf{Q}}^\alpha[t] \in \mathbb{R}^N$ denotes the predicted α -quantiles at time step t across the full collection. The framework accommodates additional inputs at the encoding block (e.g. conditioning on $\mathbf{X}[<t]$ and $\mathbf{U}[<t]$). The model is trained by minimising the pinball loss for each α -quantile at each time step in the calibration set across all N time series. Through the message-passing layers, the residuals of each time series contribute to estimating the quantiles of the error distribution at neighbouring nodes. In practice, the regressor takes as input only the most recent window of observations rather than the full sequence, and the window length, W , may differ from that used by the point predictor.

Constructing the prediction intervals. Given the trained quantile network, $\mathcal{Q}_\Psi$, prediction intervals for each target time step are constructed as

$$\hat{C}_{i,t}^\alpha = \left[\hat{x}_i[t+H] + \hat{q}_i^{\alpha/2}[t], \hat{x}_i[t+H] + \hat{q}_i^{1-\alpha/2}[t] \right], \quad (6.10)$$

or, alternatively, by optimising the interval width through a shift parameter,

$$\hat{\beta}_i = \arg \min_{\beta_i} \left| \hat{q}_i^{1-\alpha/2+\beta_i}[t] - \hat{q}_i^{\alpha/2+\beta_i}[t] \right|, \quad (6.11)$$

$$\hat{C}_{i,t}^\alpha = \left[\hat{x}_i[t+H] + \hat{q}_i^{\alpha/2+\hat{\beta}_i}[t], \hat{x}_i[t+H] + \hat{q}_i^{1-\alpha/2+\hat{\beta}_i}[t] \right]. \quad (6.12)$$

Both constructions correspond to the same confidence level, but the optimised interval in (6.12) can be narrower at the expense of the additional computation required to obtain $\hat{\beta}_i$ [C. Xu and Y. Xie, 2023a]. The choice between the two can be guided by computational constraints and the performance difference observed on a validation set. It is noted that $\hat{C}_{i,t}^\alpha$ provides only approximate coverage, as it is subject to the approximation error of the learned quantile function; Sec. 6.3.3

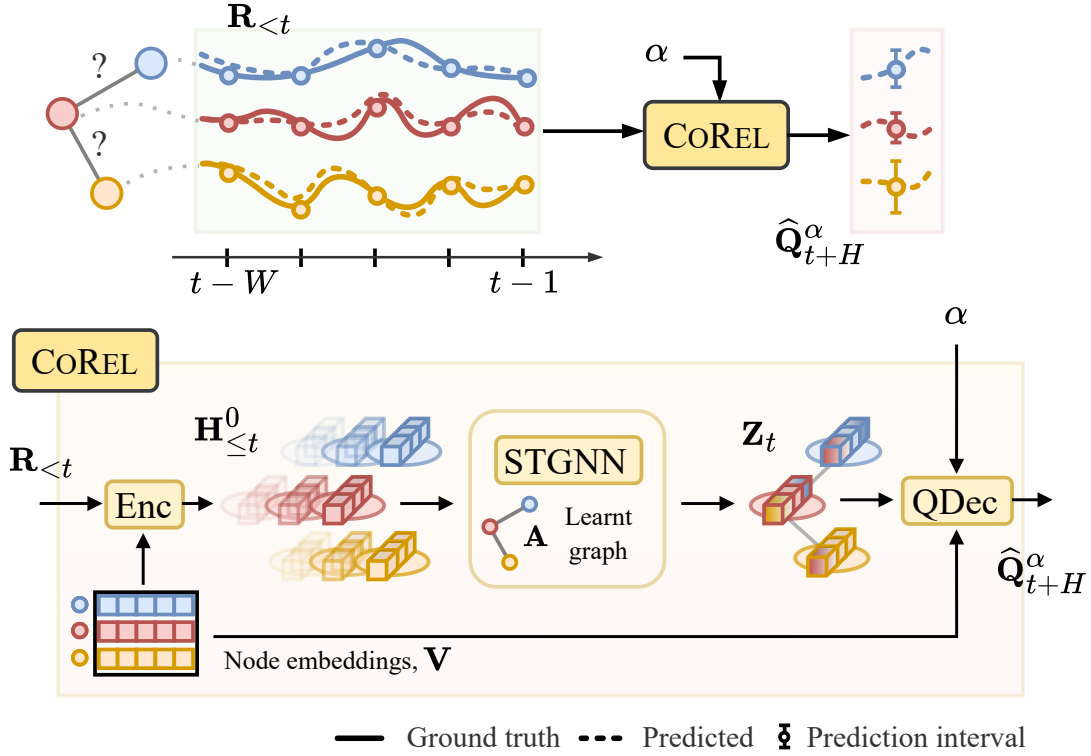


Figure 6.1: Overview of the COREL framework. Prediction residuals from a pre-trained point forecaster are fed into a hybrid global-local quantile network (Sec. 6.3.1) that follows the time-then-space STGNN template of Sec. 2.3.2. A temporal encoder (single-layer GRU) processes each node’s residual history independently, after which message-passing layers propagate information across a sparse learned graph (Sec. 6.3.2). The quantile decoder maps the resulting spatiotemporal representations, combined with learnable node embeddings, $\mathbf{v}_i$, to predicted quantiles of the residual distribution at each node, from which prediction intervals are constructed.

analyses this aspect.

A potential limitation of COREL is that residuals cannot be assumed exchangeable in most practical scenarios. The error distribution may be non-stationary, making it difficult to obtain formal coverage guarantees. To mitigate this issue, Sec. 6.3.4 introduces an efficient and scalable approach to make the framework adaptive by updating the local components (node embeddings) of the architecture over time. Finally, the use of relational components in COREL relies on the actual presence of inter-series dependencies in the data. In practical applications, the presence of spatiotemporal correlations in the residuals can be verified through statistical tests [Zambon and Alippi, 2022; Zambon and Alippi, 2023], whose outcome can support the decision to adopt the relational approach.

6.3.2 Learning the Relational Structure

When the dependency structure across time series is unknown, a graph learning module is integrated into the architecture to derive the operational graph topology directly from the residuals. The probabilistic structure learning framework of Cini, Zambon, and Alippi (2023) is adopted, which learns a distribution over sparse graphs and samples discrete adjacency matrices that can be used directly in the message-passing layers of (6.7). The key idea is to associate each potential directed edge, $j \rightarrow i$, with a learnable score, ϕ_{ij} , and to parametrise a distribution over K -nearest-neighbour graphs via the score matrix, $\Phi \in \mathbb{R}^{N \times N}$ [Cini, Zambon, and Alippi, 2023; Kazi et al., 2022; Niculae et al., 2023].

Concretely, for each node i , the graph learning module samples a neighbourhood of K nodes without replacement from a categorical distribution over the candidate set $\{1, \dots, N\}$. The score matrix is first computed as

$$\Phi = \mathcal{E}_{\xi}(\mathbf{R}[\leq t], \mathbf{V}, \dots), \quad (6.13)$$

where $\mathcal{E}_{\xi}(\cdot)$ is a trainable encoder with parameters ξ . Each row $\phi_{i,:}$ of Φ defines a categorical distribution over candidate neighbours for node i , with probabilities proportional to the exponentiated scores,

$$p(j | i) = \frac{\exp\{\phi_{ij}\}}{\sum_{k=1}^N \exp\{\phi_{ik}\}}, \quad j \in \{1, \dots, N\}. \quad (6.14)$$

A subset $\mathcal{S}_i \subset \{1, \dots, N\}$ of K neighbours is then sampled without replacement from this distribution and the i -th row of the binary adjacency matrix is constructed as

$$A_{ij} = \begin{cases} 1, & \text{if } j \in \mathcal{S}_i, \\ 0, & \text{otherwise,} \end{cases} \quad (6.15)$$

yielding a directed K -nearest-neighbour graph with at most NK edges. The resulting adjacency matrix, $\mathbf{A}$, is binary and sparse by construction, ensuring that the subsequent message-passing operations in (6.7) scale as $\mathcal{O}(NK)$ rather than $\mathcal{O}(N^2)$.

In the general formulation of (6.13), the encoder $\mathcal{E}_{\xi}$ can condition the scores on the observed residuals and node embeddings, enabling a dynamic graph distribution that adapts to the input at each time step. In the static case, the scores are learnable parameters directly, $\Phi = \xi$, removing the dependence on $\mathbf{R}[\leq t]$ and yielding a single graph distribution shared across all time steps. The static parametrisation is adopted in this chapter, as the relational structure among prediction residuals is assumed to be stable over the calibration period. The extension to dynamic, input-conditioned graphs is discussed by Cini, Zambon, and Alippi (2023) but remains an open direction

for future work within the conformal prediction setting.

End-to-end learning. Sampling without replacement from a categorical distribution is a discrete operation that precludes standard backpropagation. In practice, sampling is performed efficiently using the GumbelTopK trick [Kool, Van Hoof, and Welling, 2019], and gradients are propagated through the discrete sampling step using the continuous relaxations of S. M. Xie and Ermon (2019) paired with a straight-through gradient estimator [Bengio, Léonard, and Courville, 2013]. The gradients are sparsified by backpropagating only through the sampled edges for each node plus 10% of the remaining zero entries of the adjacency matrix, reducing computational cost while maintaining sufficient gradient signal. To allow the model to sample fewer than K neighbours when appropriate, a set of dummy nodes is introduced and discarded from the sampled graph before any message-passing operation [Cini, Zambon, and Alippi, 2023]. This mechanism enables an adaptive effective degree per node, so that the sparsity constraint K acts as an upper bound rather than a fixed neighbourhood size. Different parametrisations and gradient estimators for subset samplers can be considered, and if the sparsity assumption is deemed unrealistic for a particular problem, alternative distributions (e.g. based on independent Bernoulli random variables per edge) may be used [Cini, Zambon, and Alippi, 2023]. The complete architecture is illustrated in Fig. 6.1.

6.3.3 Theoretical Analysis

Since the prediction intervals constructed by COREL rely on a learned approximation of the residual quantile function rather than exact empirical quantiles, they provide only approximate coverage. The following proposition formalises this by bounding the coverage gap in terms of the approximation quality of the learned quantile network, $\mathcal{Q}_\Psi$ (6.9).

Proposition 6.3.1. *Let $P_{t+H}^c(x_i[t+H]) = p_{t+H}(x_i[t+H] \mid \mathbf{X}[<t], \mathbf{U}[<t])$ denote the true conditional data-generating distribution at time step $t+H$, let $P_\Psi^c(x_i[t+H]) = p_\Psi(x_i[t+H] \mid \mathbf{X}[<t], \mathbf{U}[<t])$ denote the distribution implied by the learned quantile function, $\mathcal{Q}_\Psi$, and let $\widehat{C}_{i,t}^\alpha$ denote the prediction interval at node i and time step t for miscoverage level α , constructed via either (6.10) or (6.12). Then*

$$P_{t+H}^c\left(x_i[t+H] \in \widehat{C}_{i,t}^\alpha\right) \geq 1 - \alpha - \text{TV}\left(P_\Psi^c, P_{t+H}^c\right), \quad (6.16)$$

where $\text{TV}(P, Q) := \sup_B |P(B) - Q(B)|$ denotes the total variation distance between distributions P and Q .

Proof. From the definition of total variation distance, it follows that for any event B ,

$$\begin{aligned} |P_{t+H}^c(B) - P_{\Psi}^c(B)| &\leq \text{TV}(P_{\Psi}^c, P_{t+H}^c), \\ \implies P_{t+H}^c(B) &\geq P_{\Psi}^c(B) - \text{TV}(P_{\Psi}^c, P_{t+H}^c). \end{aligned} \quad (6.17)$$

For $\hat{C}_{i,t}^{\alpha}$ constructed via either (6.10) or (6.12), it holds by construction that $P_{\Psi}^c(x_i[t+H] \in \hat{C}_{i,t}^{\alpha}) = 1 - \alpha$. Substituting into (6.17) with $B = \{x_i[t+H] \in \hat{C}_{i,t}^{\alpha}\}$ yields the result. $\square$

Importantly, this result does not require the data-generating process to be time-invariant. [Prop. 6.3.1](#) links the conditional coverage gap to the approximation error in estimating the quantile function of the residuals, sharing structural similarities with the result of R. F. Barber et al. (2023) that bounds the miscoverage gap for conformal prediction from weighted empirical quantiles. [Prop. 6.3.1](#) can be viewed as an analogous result where empirical quantile estimates are replaced by a parametric function approximation. Under assumptions on the expressivity of the quantile regressor in (6.9) and on the mixing properties of the stochastic process (e.g. strong mixing), the total variation between the learned and true distributions can be expected to shrink asymptotically as the calibration set grows. Moreover, in this regime, monitoring the coverage gap on a validation set provides an estimate of the miscoverage on test data. Similar analyses have been carried out for related conformal prediction methods [C. Xu and Y. Xie, 2023b; Jonghyeok Lee, C. Xu, and Y. Xie, 2025].

If the data-generating process is non-stationary, the quantile network, $\mathcal{Q}_{\Psi}$, must be updated over time to keep the coverage gap contained. The following section introduces a sample-efficient approach to achieve this. Regarding computational complexity, the cost of a forward pass scales as $\mathcal{O}(WN + |\mathcal{E}|)$ for a TTS architecture with window size W and $|\mathcal{E}|$ edges [Cini, Marisca, Zambon, et al., 2025], maintaining the scalability properties discussed in [Sec. 2.3.2](#).

6.3.4 Adaptation

The relational quantile network introduced in [Sec. 6.3.1](#) can yield arbitrarily large coverage gaps in the presence of distribution shifts from the calibration set. Adopting a full re-training approach, as in C. Xu and Y. Xie (2023b), would be impractical due to the sample complexity entailed by the deep learning architecture. To mitigate this issue while keeping computational cost under control, only the local components of the model are updated over time, specifically, the learnable node embeddings, $\mathbf{V}$ [Cini, Marisca, Zambon, et al., 2024]. This allows most learnable parameters to remain fixed while fine-tuning only a small number of weights per node. This approach connects to the adaptive filtering principles developed in [Sec. 2.1.4](#): just as the LMS algorithm updates

filter coefficients incrementally to track non-stationary environments, the embedding update here provides a lightweight adaptation mechanism that tracks changes in the error distribution over time.

6.4 Experiments

COREL is validated across three experimental settings: (i) a comparison against state-of-the-art conformal prediction methods on a range of forecasting benchmarks (Sec. 6.4.1), (ii) an analysis in a controlled synthetic environment where the data-generating process is known (Sec. 6.4.2), and (iii) an assessment of the adaptive embedding update procedure (Sec. 6.4.3).

COREL is implemented as a recurrent neural network followed by two message-passing layers, adopting the TTS template of Sec. 2.3.2. To approximate the quantile function, the model is trained by minimising the pinball loss function over a discrete set of quantiles [R. Wen et al., 2017]. Prediction intervals are constructed using (6.10); results for the optimised construction in (6.12) are reported in Sec. 6.4.4. The graph is learned by directly parametrising the score matrix, Φ , associating a learnable parameter with each entry, and using a fixed neighbourhood size of $K = 20$ for the graph learning module.

Evaluation metrics. Prediction intervals are evaluated using three complementary metrics at miscoverage level α . Given a prediction interval

$$\widehat{C}_{i,t}^\alpha = [\hat{x}_i[t+H] + \hat{q}_i^{\alpha/2}[t], \hat{x}_i[t+H] + \hat{q}_i^{1-\alpha/2}[t]]$$

and the corresponding true value, $x_i[t+H]$, the coverage gap measures the difference between achieved and nominal coverage as

$$\Delta\text{Cov}_i = 100 \left(\mathbb{1} \left(x_i[t+H] \in \widehat{C}_{i,t}^\alpha \right) - (1 - \alpha) \right), \quad (6.18)$$

where $\mathbb{1}(\cdot)$ is the indicator function; a value of zero indicates exact calibration. The prediction interval width measures the size of the interval directly as

$$\text{PI-Width}_i = \hat{q}_i^{1-\alpha/2}[t] - \hat{q}_i^{\alpha/2}[t]. \quad (6.19)$$

Since coverage and width are competing objectives (a trivially wide interval achieves perfect coverage but conveys no useful information), the Winkler score [Winkler, 1972] is additionally reported, which equals the interval width when the true value falls within the interval and adds a

penalty proportional to the magnitude of the miss otherwise,

$$W_i = \begin{cases} \text{PI-Width}_i + \frac{2}{\alpha} \left(\hat{q}_i^{\alpha/2}[t] - x_i[t+H] \right), & \text{if } x_i[t+H] < \hat{q}_i^{\alpha/2}[t], \\ \text{PI-Width}_i, & \text{if } x_i[t+H] \in \hat{C}_{i,t}^\alpha, \\ \text{PI-Width}_i + \frac{2}{\alpha} \left(x_i[t+H] - \hat{q}_i^{1-\alpha/2}[t] \right), & \text{if } x_i[t+H] > \hat{q}_i^{1-\alpha/2}[t]. \end{cases} \quad (6.20)$$

All three metrics are averaged over nodes and time steps.

Implementation details. All experiments were implemented in Python [Van Rossum and Drake, 2009] using PyTorch [Paszke et al., 2019], PyTorch Lightning [Falcon and The PyTorch Lightning team, 2019], PyTorch Geometric [Fey and Lenssen, 2019], and Torch Spatiotemporal [Cini and Marisca, 2022]. Experiments were conducted on a server equipped with AMD EPYC 7513 CPUs and NVIDIA RTX A5000 GPUs. For the HopCPT, SPCI, NexCP, and SeqCP baselines, the open-source implementation of Auer et al. (2023) is used.

6.4.1 Time Series Forecasting Benchmarks

Three datasets from different application domains are considered. The **METR-LA** traffic dataset [Yaguang Li et al., 2018] consists of 34,272 time steps of measurements from 207 loop detectors sampled at 5-minute intervals on the Los Angeles County highway network; a window of $W = 12$ time steps is used and the 12-step-ahead observation is predicted. The **AQI** dataset [Zheng et al., 2015] comprises 8,760 hourly PM2.5 measurements from 437 monitoring stations across China; $W = 24$ is used and 3 steps ahead are predicted. The **CER-E** dataset [Commission for Energy Regulation, 2016; Cini, Marisca, and Alippi, 2022] contains 25,728 time steps of energy consumption readings aggregated at 30-minute intervals from 485 smart meters monitoring small and medium-sized enterprises; $W = 48$ is used and 5 steps ahead are predicted. Established preprocessing steps [Yaguang Li et al., 2018; Z. Wu et al., 2019; Cini, Marisca, Zambon, et al., 2024] are followed and 40%/40%/20% splits are adopted for training, calibration, and testing, with the first 25% of the calibration data reserved as a validation set for early stopping. Input features are scaled using standard scaling across time series.

For each dataset, three baseline point predictors are trained: a simple RNN with GRU cells [Cho, Merriënboer, et al., 2014] (1 layer, hidden size 32), a decoder-only Transformer [Vaswani et al., 2017] (hidden size 32, feed-forward size 64, 2 attention heads, 3 layers, dropout 0.1), and a TTS STGNN following the template in Sec. 2.3.2 (hidden size 32, node embedding size 16, 1-layer GRU, 2 message-passing layers). All base models are trained by minimising the MAE loss us-

ing the Adam optimiser [D. P. Kingma and Ba, 2015] for 200 epochs with batch size 32. The STGNN uses a pre-defined graph encoding the inter-series dependencies: for METR-LA and AQI, the adjacency matrix is constructed from a thresholded Gaussian kernel on pairwise geographic distances [Shuman et al., 2013]; for CER-E, a 10-nearest-neighbour graph is extracted from week-wise correntropy similarities [Cini, Marisca, and Alippi, 2022]. After training, each baseline is evaluated on the calibration set and the associated residuals are saved for use by the conformal prediction methods.

Baselines. COREL is compared against five conformal prediction methods. (i) SCP, the standard approach with a single global quantile. (ii) Sequential conformal prediction (SeqCP) [C. Xu and Y. Xie, 2023a], which computes empirical quantiles from only the most recent K residuals, with the window size K tuned over values from 10 to 200. (iii) Non-exchangeable conformal prediction (NexCP) [R. F. Barber et al., 2023], which assigns exponentially decaying weights to past residuals, with the decay parameter, ρ , tuned over values from 0.9 to 0.999. (iv) Sequential conformal prediction intervals (SPCI) [C. Xu and Y. Xie, 2023b], which estimates the residual quantile function using a quantile random forest on the most recent 100 observations, with separate models trained for each combination of node and target coverage level. (v) HopCPT [Auer et al., 2023], which reweights past residuals using attention scores from a Modern Hopfield Network [Ramsauer et al., 2021], trained for 3,000 epochs with batch size 4, following the model selection procedure of Auer et al. (2023). These baselines are representative of the two main recent paradigms: quantile regression (SPCI) and residual reweighting (HopCPT). An ablation, CORNN, is also included, which uses the same architecture as COREL but removes the message-passing layers and node embeddings.

For COREL and CORNN, the number of hidden units in the STGNN is tuned via a small grid search on 10% of the calibration data. The model is trained for a maximum of 100 epochs on the calibration set, with each epoch consisting of at most 50 mini-batches of size 64. The Adam optimiser [D. P. Kingma and Ba, 2015] is used with an initial learning rate of 0.003, reduced by 75% every 20 epochs. Except for HopCPT, which uses a custom procedure [Auer et al., 2023], model selection is performed on a validation set by optimising the Winkler score.

Results. Tab. 6.1 reports results across datasets and base prediction models. COREL outperforms the competitors in Winkler score in nearly all configurations. Exceptions arise primarily when the base model is itself a STGNN with access to a pre-defined graph, as it already captures some spatiotemporal dependencies. However, pre-defined graphs are not always available in practical applications. In terms of coverage, COREL achieves good results, with some exceptions in

Table 6.1: Uncertainty quantification performance for $\alpha = 0.1$ across three datasets and three base predictors. ΔCov denotes the coverage gap (actual minus nominal coverage; closer to zero is better), PI-Width the mean prediction interval width ($\downarrow$), and Winkler the Winkler score ($\downarrow$), which jointly penalises miscoverage and interval width. ΔCov values are colour-coded: **green** (0–2%), **yellow** (2–3%), **orange** (3–4%), **red** (>4%). Best Winkler score for each scenario in **bold**. Results for CoRNN and CoREL reported as mean $\pm$ standard deviation across 3 random seeds.

	Metric	SCP	NexCP	SeqCP	SPCI	HopCPT	CoRNN	CoREL
<i>METR-LA</i>								
RNN	ΔCov	-1.28	-1.00	-6.93	-1.24 $\pm$ 0.01	-1.46 $\pm$ 0.11	-0.49 $\pm$ 0.53	-1.29 $\pm$ 0.21
	PI-Width	20.69	25.23	19.52	19.84 $\pm$ 0.01	16.62 $\pm$ 0.10	19.48 $\pm$ 0.43	14.38 $\pm$ 0.22
	Winkler	40.80	41.33	50.12	37.95 $\pm$ 0.01	25.63 $\pm$ 0.19	30.24 $\pm$ 0.18	23.78$\pm$0.20
TRANSF	ΔCov	-1.18	-0.97	-6.98	-1.13 $\pm$ 0.00	-1.22 $\pm$ 0.56	-0.55 $\pm$ 0.61	-1.02 $\pm$ 0.63
	PI-Width	20.81	25.30	19.44	19.86 $\pm$ 0.01	16.66 $\pm$ 0.13	19.38 $\pm$ 0.64	14.17 $\pm$ 0.43
	Winkler	40.55	41.64	49.90	37.74 $\pm$ 0.01	25.37 $\pm$ 0.33	30.33 $\pm$ 0.18	23.65$\pm$0.19
STGNN	ΔCov	-0.99	-0.62	-13.60	-0.87 $\pm$ 0.00	-0.13 $\pm$ 0.27	-0.34 $\pm$ 0.49	-0.92 $\pm$ 0.31
	PI-Width	17.30	22.20	12.87	16.38 $\pm$ 0.01	15.75 $\pm$ 0.19	16.20 $\pm$ 0.24	14.65 $\pm$ 0.24
	Winkler	34.94	34.49	40.36	33.66 $\pm$ 0.01	22.80$\pm$0.28	28.74 $\pm$ 0.12	24.70 $\pm$ 0.16
<i>CER-E</i>								
RNN	ΔCov	-3.46	0.11	-3.57	-3.45 $\pm$ 0.00	-4.37 $\pm$ 0.32	-2.24 $\pm$ 0.34	-3.60 $\pm$ 0.35
	PI-Width	2.60	3.26	2.75	2.39 $\pm$ 0.00	1.97 $\pm$ 0.03	1.96 $\pm$ 0.02	1.83 $\pm$ 0.03
	Winkler	5.69	5.48	5.79	5.29 $\pm$ 0.00	3.87 $\pm$ 0.07	3.84 $\pm$ 0.01	3.71$\pm$0.04
TRANSF	ΔCov	-3.35	0.10	-3.53	-3.26 $\pm$ 0.01	-3.95 $\pm$ 0.60	-2.04 $\pm$ 0.32	-3.97 $\pm$ 0.24
	PI-Width	2.52	3.16	2.67	2.33 $\pm$ 0.00	2.01 $\pm$ 0.09	1.94 $\pm$ 0.03	1.80 $\pm$ 0.02
	Winkler	5.60	5.36	5.69	5.20 $\pm$ 0.00	3.88 $\pm$ 0.12	3.82 $\pm$ 0.02	3.67$\pm$0.02
STGNN	ΔCov	-4.30	0.08	-3.83	-4.17 $\pm$ 0.01	-5.06 $\pm$ 0.15	-2.13 $\pm$ 0.83	-4.99 $\pm$ 0.61
	PI-Width	2.28	3.00	2.42	2.09 $\pm$ 0.00	1.79 $\pm$ 0.01	1.85 $\pm$ 0.04	1.77 $\pm$ 0.05
	Winkler	5.11	4.87	4.99	4.76 $\pm$ 0.00	3.49$\pm$0.02	3.72 $\pm$ 0.02	3.76 $\pm$ 0.04
<i>AQI</i>								
RNN	ΔCov	5.09	-0.63	-3.21	1.79 $\pm$ 0.01	0.01 $\pm$ 2.62	-2.06 $\pm$ 1.52	-2.78 $\pm$ 0.60
	PI-Width	118.06	82.04	74.73	103.88 $\pm$ 0.03	90.35 $\pm$ 13.04	71.61 $\pm$ 1.82	68.13 $\pm$ 1.30
	Winkler	148.61	131.18	135.59	143.10 $\pm$ 0.01	133.24 $\pm$ 8.31	113.11 $\pm$ 0.77	107.67$\pm$0.94
TRANSF	ΔCov	5.07	-0.66	-3.17	1.60 $\pm$ 0.01	-2.34 $\pm$ 1.19	-1.93 $\pm$ 0.44	-1.81 $\pm$ 1.67
	PI-Width	118.50	81.15	75.08	104.11 $\pm$ 0.01	80.37 $\pm$ 2.09	73.25 $\pm$ 0.95	72.25 $\pm$ 3.16
	Winkler	150.33	132.08	137.62	145.14 $\pm$ 0.02	129.85 $\pm$ 3.14	112.71 $\pm$ 0.44	108.71$\pm$1.38
STGNN	ΔCov	4.48	-0.32	-2.94	2.64 $\pm$ 0.01	-0.78 $\pm$ 1.34	-1.52 $\pm$ 0.51	-2.00 $\pm$ 1.54
	PI-Width	111.68	80.01	71.92	99.90 $\pm$ 0.03	79.39 $\pm$ 3.38	70.17 $\pm$ 0.80	68.24 $\pm$ 2.65
	Winkler	143.39	127.98	130.86	137.09 $\pm$ 0.03	121.26 $\pm$ 2.48	109.65 $\pm$ 0.30	108.45$\pm$1.52

Table 6.2: Uncertainty quantification performance on the synthetic GPVAR dataset for $\alpha = 0.1$. Two base predictors are evaluated: an RNN (no graph structure) and an STGNN (with ground-truth graph). ΔCov denotes the coverage gap (%), PI-Width the mean prediction interval width ($\downarrow$), and Winkler the Winkler score ($\downarrow$). The variant “w/ true $\mathbf{A}$ ” provides COREL with the ground-truth adjacency matrix instead of learning it. Best Winkler score for each predictor in **bold**. Results for CORNN and COREL reported as mean $\pm$ standard deviation across 3 random seeds.

Model	GPVAR-RNN			GPVAR-STGNN		
	ΔCov	PI-Width	Winkler	ΔCov	PI-Width	Winkler
SCP	-0.02	1.67	2.14	-0.02	1.32	1.66
CoRNN	-0.1 $\pm$ 0.2	1.63 $\pm$ 0.01	2.04 $\pm$ 0.00	0.0 $\pm$ 0.0	1.32 $\pm$ 0.00	1.66 $\pm$ 0.00
COREL	0.0 $\pm$ 0.1	1.33 $\pm$ 0.00	1.67$\pm$0.00	0.1 $\pm$ 0.1	1.32 $\pm$ 0.01	1.66$\pm$0.00
w/ true $\mathbf{A}$	-0.1 $\pm$ 0.1	1.32 $\pm$ 0.00	1.66$\pm$0.00	0.0 $\pm$ 0.2	1.32 $\pm$ 0.01	1.66$\pm$0.00

the CER-E dataset where model selection prioritised the Winkler score over coverage alone. The ablation CoRNN obtains competitive overall performance against the state-of-the-art baselines but is outperformed by COREL in most settings in terms of Winkler score, confirming the value of the relational component. Among the competitors, HopCPT provides competitive results in METR-LA and CER-E but with a larger coverage gap. Except for SeqCP, the other baselines obtain good coverage at the expense of substantially wider prediction intervals.

Regarding computational scalability, COREL shares most learnable parameters among time series, enabling efficient parallelisation on a GPU. Beyond the learnable node embeddings, $\mathbf{V}$, the method relies only on a short window of recent observations as input at each time step. By contrast, SPCI requires training a separate model for each time series, and HopCPT requires computing attention scores over the entire calibration set at each step. To illustrate the practical difference, on the METR-LA dataset (the smallest in terms of number of time series), training and testing require approximately 3 days for SPCI, approximately 11 hours for HopCPT, and approximately 5 minutes for COREL.

6.4.2 Controlled Synthetic Environment

COREL is first evaluated in a controlled setting where the data-generating process and ground-truth graph are known, using the GPVAR benchmark [Zambon and Alippi, 2022] with a setup analogous to Cini, Marisca, Zambon, et al. (2024). Data are generated recursively from a polynomial graph

filter applied to the adjacency matrix, $\mathbf{A}$,

$$\begin{aligned}\mathbf{H}[t] &= \sum_{p=1}^P \sum_{q=1}^Q \Theta_{q,p} \mathbf{A}^{p-1} \mathbf{X}[t-q], \\ \mathbf{X}[t+1] &= \mathbf{a} \odot \tanh(\mathbf{H}[t]) + \mathbf{b} \odot \tanh(\mathbf{X}[t-1]) + \boldsymbol{\eta}[t],\end{aligned}\tag{6.21}$$

where $\boldsymbol{\Theta} \in \mathbb{R}^{Q \times K}$ collects the filter coefficients, $a, b \in \mathbb{R}$ are scalar mixing weights (broadcast across nodes), and $\boldsymbol{\eta}[t] \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ is additive Gaussian noise. This process is a non-linear extension of the graph VAR model introduced in [Sec. 2.3.1](#). The simulation uses 40,000 time steps on an undirected community graph with 60 nodes, with process parameters

$$\boldsymbol{\Theta} = \begin{bmatrix} 2.5 & -2.0 & -0.5 \\ 1.0 & 3.0 & 0.0 \end{bmatrix}, \quad a = b = 0.5, \quad \sigma = 0.4.\tag{6.22}$$

An input window of $W = 5$ time steps is used to predict the next observation. For COREL, the number of hidden units is fixed to 16 per layer with an embedding size of 8. Both a recurrent neural network and a STGNN are used as base point predictors. As shown by Zambon and Alippi (2022), STGNNs achieve forecasting accuracy near the theoretical optimum in this dataset, yielding approximately uncorrelated residuals. Standard split conformal prediction should therefore suffice in this case. The objective of this experiment is to verify that COREL matches the performance of standard conformal prediction when residual structure is minimal, while demonstrating that it can exploit spatiotemporal dependencies when they are present (as with the weaker RNN predictor, whose residuals retain graph-structured correlations).

Results. [Tab. 6.2](#) reports the results. When using a recurrent neural network as the point predictor (GPVAR-RNN), COREL significantly outperforms both standard split conformal prediction and the CORNN ablation. Furthermore, COREL achieves results closely matching those obtained with direct access to the ground-truth graph, demonstrating the effectiveness of the learned graph in capturing latent relational dependencies. [Fig. 6.2](#) provides a qualitative comparison between the true graph and the graph learned by COREL, obtained by extracting the top- K scores from each row of the score matrix, $\boldsymbol{\Phi}$. The learned graph recovers the true edges while including additional links; importantly, perfectly recovering the ground-truth topology is not required to achieve near-optimal performance. Given the injected Gaussian noise with $\sigma = 0.4$, the theoretical optimum interval width for asymptotic 90% marginal coverage is 1.315. COREL achieves essentially perfect coverage with interval width close to this theoretical limit, while CORNN requires a substantially wider interval to obtain similar coverage. When using a spatiotemporal graph neural network as

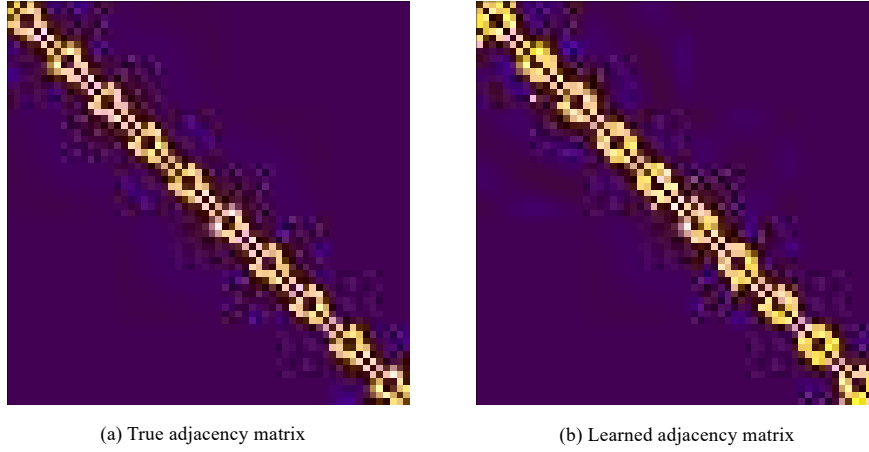


Figure 6.2: Comparison between (a) the true graph and (b) the graph structure learned by COREL in GPVAR.

Table 6.3: Effect of adaptive embedding updates on COREL performance for the CER-E dataset ($\alpha = 0.1$). The adaptive variant updates node embeddings incrementally over six folds of the test set (Sec. 6.3.4). ΔCov denotes the coverage gap (%), PI-Width the mean prediction interval width ($\downarrow$), and Winkler the Winkler score ($\downarrow$). ΔCov values are colour-coded: **yellow** (2–3%), **orange** (3–4%), **red** ($>4\%$). Best Winkler score for each predictor in **bold**. Results reported as mean $\pm$ standard deviation across 3 random seeds.

Predictor	Adaptation	ΔCov	PI-Width	Winkler
RNN	W/o adapt.	−3.44±0.40	1.84±0.02	3.71±0.03
	W/ adapt.	−2.70±0.27	1.85±0.01	3.49±0.02
TRANSF	W/o adapt.	−3.08±0.15	1.84±0.02	3.69±0.04
	W/ adapt.	−2.42±0.20	1.85±0.01	3.49±0.04
STGNN	W/o adapt.	−4.07±0.47	1.80±0.03	3.78±0.03
	W/ adapt.	−3.06±0.25	1.81±0.02	3.58±0.02

the base model (GPVAR-STGNN), standard split conformal prediction is sufficient, confirming that the base model already captures the relevant dependencies, as expected.

6.4.3 Adaptation Experiments

This experiment evaluates the effectiveness of the adaptive embedding update proposed in Sec. 6.3.4. The focus is on the CER-E dataset, where the calibration set does not cover a full year, likely introducing a distribution shift at test time due to the seasonality of energy consumption. COREL is configured with 64 hidden units in each encoder and decoder layer and an embedding size of 32, trained on the full calibration set with a fixed hyperparameter configuration across scenarios. At test time, the node embeddings are updated every M time steps, where M corresponds to $\frac{1}{6}$ of the test set length. In practice, the test set is split into $K = 6$ folds and the model is iteratively

Table 6.4: Effect of prediction interval width optimisation via the quantile offset $\hat{\beta}_i$ in (6.11) on coverage and interval width ($\alpha = 0.1$). For each dataset, predictor, and method, the first row reports the symmetric interval of (6.10) and the second row the optimised interval of (6.12). The optimised interval consistently reduces PI-Width at the cost of a slight increase in the coverage gap. Results reported as mean $\pm$ standard deviation across 3 random seeds.

	RNN		TRANSF		STGNN	
	ΔCov	PI-Width	ΔCov	PI-Width	ΔCov	PI-Width
<i>METR-LA</i>						
CoRNN	-0.70 ± 0.20	19.69 ± 0.27	-0.52 ± 0.56	19.07 ± 0.42	-0.67 ± 0.47	16.02 ± 0.13
$\hat{\beta}$ optim.	-0.80 ± 0.28	17.71 ± 0.28	-0.96 ± 0.35	17.16 ± 0.46	-0.88 ± 0.44	14.98 ± 0.17
CoREL	-1.25 ± 0.80	14.24 ± 0.53	-0.97 ± 0.51	14.43 ± 0.38	-1.31 ± 0.48	14.36 ± 0.20
$\hat{\beta}$ optim.	-1.49 ± 0.56	13.23 ± 0.46	-1.26 ± 0.42	13.47 ± 0.35	-1.44 ± 0.45	13.46 ± 0.20
<i>CER-E</i>						
CoRNN	-2.09 ± 0.56	1.93 ± 0.01	-1.70 ± 0.37	1.99 ± 0.02	-2.45 ± 0.41	1.85 ± 0.03
$\hat{\beta}$ optim.	-2.63 ± 0.65	1.82 ± 0.01	-2.28 ± 0.31	1.88 ± 0.02	-3.10 ± 0.52	1.78 ± 0.03
CoREL	-3.86 ± 0.53	1.83 ± 0.02	-3.81 ± 0.30	1.83 ± 0.02	-4.87 ± 0.51	1.79 ± 0.04
$\hat{\beta}$ optim.	-4.16 ± 0.59	1.75 ± 0.02	-4.10 ± 0.21	1.75 ± 0.02	-5.13 ± 0.51	1.72 ± 0.03
<i>AQI</i>						
CoRNN	-1.45 ± 0.61	73.88 ± 1.17	-1.08 ± 0.22	74.02 ± 0.74	-1.50 ± 0.63	70.42 ± 0.94
$\hat{\beta}$ optim.	-1.90 ± 0.54	70.17 ± 1.22	-1.10 ± 0.25	70.95 ± 0.68	-1.38 ± 0.62	68.24 ± 0.85
CoREL	-2.40 ± 1.97	70.44 ± 2.69	-3.54 ± 1.80	68.31 ± 4.49	-2.61 ± 1.92	67.76 ± 3.12
$\hat{\beta}$ optim.	-2.83 ± 1.86	67.35 ± 2.31	-4.19 ± 1.96	65.20 ± 4.20	-3.22 ± 2.45	65.32 ± 3.43

fine-tuned on each fold by running 25 epochs of at most 10 batches each, using samples from the most recent M time steps with a fixed learning rate of 0.001; all shared parameters except the embeddings remain frozen. This procedure simulates a real-world scenario where new data arrive over time and are used for lightweight model adaptation. Results, reported in Tab. 6.3, show that this adaptation scheme consistently reduces the coverage gap across all three base predictors (e.g. from -3.44% to -2.70% for the RNN predictor) while maintaining comparable interval widths, yielding improved Winkler scores.

6.4.4 Prediction Interval Width Optimisation

As discussed in Sec. 6.3.1, the width of the prediction interval can be reduced by searching for an appropriate quantile offset $\hat{\beta}_i$ via the optimisation in (6.11), which identifies pairs of quantiles yielding the same coverage $1 - \alpha$ while achieving the smallest interval width. This is particularly relevant when the residual distribution is asymmetric. Tab. 6.4 summarises the results: the procedure consistently finds intervals with smaller width at the cost of a slight reduction in coverage, for both CoREL and the CoRNN ablation.

6.5 Discussion

Calibrated uncertainty quantification is essential for any spatiotemporal prediction system deployed in consequential settings, yet existing conformal prediction methods for time series discard the relational structure that governs how prediction errors co-vary across space. This chapter introduced COREL to address this gap, exploiting inter-series dependencies through graph-based operators to produce tighter, better-calibrated prediction intervals. The following discussion examines the sources of improvement, the pathway toward clinical application in the imputation setting, and remaining limitations.

Core validation. The experimental results isolate the contribution of each component. The CORNN ablation, which removes the message-passing layers and node embeddings, obtains competitive performance against state-of-the-art baselines but is consistently outperformed by COREL in Winkler score across most configurations, confirming that the relational component is the primary source of improvement. The controlled experiment on synthetic graph-VAR data further sharpens this conclusion: COREL achieves interval widths approaching the theoretical optimum for the known noise level, closely matching results obtained with direct access to the ground-truth graph, while CORNN requires substantially wider intervals to attain similar coverage. This demonstrates that the learned sparse graph successfully recovers latent relational dependencies without requiring the topology to be specified *a priori*. The adaptive embedding update provides a complementary benefit under distribution shift: on the energy consumption dataset, where the calibration set does not span a full seasonal cycle, periodic embedding updates consistently reduce the coverage gap while maintaining comparable interval widths. Together, these results confirm that the three components of COREL, relational message passing, learned sparse graph structure, and adaptive node embeddings, each address a distinct source of inefficiency in existing conformal prediction methods for correlated time series.

From forecasting to imputation. The method and experiments in this chapter operate in the forecasting setting, where the goal is to predict future values and quantify the associated uncertainty. However, COREL’s core machinery, a graph-based quantile network that conditions uncertainty estimates on neighbourhood observations over a learned sparse topology, is not specific to forecasting. Extending this framework to the imputation setting, where the goal is to reconstruct missing values at observed time steps as in FIBMAP ([Chapter 5](#)), is a natural next step. In this setting, conformal prediction intervals would quantify the reliability of reconstructed values at unobserved spatial locations, providing clinicians with calibrated confidence measures for

imputation maps.

Recent work in this direction includes conformalised matrix completion [Gui, R. Barber, and Ma, 2023], which demonstrates that distribution-free predictive intervals can be constructed for missing entries in low-rank matrices, and the work of Ke et al. (2025) who extended this to time-indexed data with both scattered and block missingness through autoregressive regularisation. However, both approaches treat calibration residuals as exchangeable and do not provide coverage guarantees under temporal dependence. While the forecasting baselines considered in this chapter [R. F. Barber et al., 2023; C. Xu and Y. Xie, 2023a; C. Xu and Y. Xie, 2023b; Auer et al., 2023] address this limitation for prediction of future values, none have been combined with an imputation-oriented conformal procedure for missing entries. Extending COREL to the imputation setting would require synthesising these two directions: the entry-wise conformal imputation framework of [Gui, R. Barber, and Ma, 2023; Ke et al., 2025] with the dependence-aware framework that COREL already partially instantiates. Specific challenges include defining conformity scores for imputed entries where the masking structure couples the locations of missing data to the error distribution, and establishing whether coverage can be conditioned on gap length, spatial location, or missingness type, beyond the marginal, entry-wise guarantees currently available. In the context of cardiac mapping, such an extension could provide calibrated prediction intervals for FIBMAP’s reconstructions, where the relational structure conditioning the uncertainty estimates could either be learned from imputation residuals as in COREL, or guided by the known anatomical connectivity of the atrial surface.

Limitations and future directions. Several directions remain for future work. Extending the framework to heterogeneous time series collections, where different variables describe different physical quantities, would broaden applicability. The non-stationary setting warrants further investigation, including the development of coverage guarantees under specific assumptions about the data-generating process. Applications to very large time series collections could benefit from combination with scalable graph-based architectures [Chiang et al., 2019; Cini, Marisca, Bianchi, et al., 2023]. Finally, while COREL provides separate prediction intervals for each time series, extending the framework to model the joint probability distribution across time series would capture additional dependencies.

6.6 Summary and Relation to Thesis

This chapter introduced COREL, a conformal prediction method that exploits relational structure among correlated time series to provide calibrated uncertainty quantification for spatiotemporal

forecasts. By training a graph-based quantile network on prediction residuals from a calibration set, COREL constructs prediction intervals that account for inter-series dependencies without requiring modification of the base forecasting model, a known graph topology, or distributional assumptions about the prediction errors. Results across traffic, air quality, and energy consumption benchmarks demonstrate that exploiting relational structure consistently produces tighter prediction intervals while maintaining accurate coverage guarantees, establishing that spatial dependencies in correlated time series are a valuable and underexploited source of information for uncertainty quantification.

This chapter directly addresses the second research question of this thesis (**RQ2**): *how can a relational graph structure between correlated time series be exploited to provide calibrated, distribution-free uncertainty quantification for spatiotemporal predictions?* The relational quantile network of COREL instantiates the time-then-space STGNN template of [Sec. 2.3.2](#) within a conformal prediction framework, applying the message-passing operators of [Fig. 2.2.5](#) to prediction residuals rather than raw observations. The sparse graph learning module ([Sec. 6.3.2](#)) provides a principled, data-driven alternative to the fixed topologies assumed throughout [Chapter 2](#), while the adaptive embedding update ([Sec. 6.3.4](#)) connects to the online adaptation principles developed in [Sec. 2.1.4](#).

Two open challenges identified in this chapter point toward the final contribution of this thesis. First, COREL’s graph is learned once from the calibration set rather than tracked over time, yet the functional connectivity governing spatiotemporal dynamics may itself evolve. Second, while the graph learning component sharpens uncertainty estimates, the learned topology is a means to that end rather than an object of study in its own right. The next chapter shifts focus entirely to dynamic graph structure learning, developing ADACGP, an online system identification method for streaming collections of time series that extends the causal graph process framework of [Sec. 2.3.1](#) to the adaptive setting. In contrast to the static, deep-learning-based graph estimation in COREL, ADACGP takes a fundamentally different and sample-efficient approach to learning the graph topology. Specifically, ADACGP tracks the evolution of directed causal structure in real time, enabling the characterisation of dynamic propagation patterns, such as the time-varying functional connectivity of cardiac fibrillation ([Sec. 3.3](#)), that a static graph cannot capture.

Chapter 7

Online Graph Topology Learning via Time-Vertex Adaptive Filters

This chapter introduces ADACGP (Adaptive identification of Causal Graph Processes), a sparsity-aware adaptive algorithm for estimating time-varying graph topologies from streaming multivariate time series. ADACGP extends the causal graph process framework of [Sec. 2.3.1](#) to the online setting, estimating the graph shift operator through recursive update formulae that addresses sparsity, shift-invariance, and estimation bias. A variable splitting approach enables reliable identification of causal connections with near-zero false alarm rates, overcoming a fundamental limitation of existing adaptive methods. Through comprehensive simulations across diverse graph topologies, ADACGP achieves improvements exceeding 83% in graph shift operator estimation compared to state-of-the-art methods while maintaining favourable computational scaling. Applied to ex vivo optical mapping recordings of rat ventricular fibrillation, ADACGP tracks dynamic changes in propagation patterns more effectively than batch Granger causality, capturing temporal variations in functional connectivity that static approaches miss. The algorithm identifies stability characteristics in conduction patterns that may maintain arrhythmias, completing the reconstruct–analyse–characterise pipeline that structures this thesis.

7.1	Introduction	157
7.2	Methodology	158
7.3	Experiments	168
7.4	Discussion	184
7.5	Summary and Relation to Thesis	186

This chapter is based on Jenkins, Varidhisai, et al. (2025).

7.1 Introduction

The graph topology, $\mathcal{G}$, has been treated throughout the theoretical background (Chapter 2) as given inputs to each model, from polynomial graph filters to STGNNs. Yet in many domains the relational structure is unknown and must be inferred from the observed signals, as identified in Sec. 2.3.3. In neuroscience, functional connectivity between brain regions is estimated from co-activation patterns and may reconfigure during cognitive tasks [Preti, Bolton, and Van De Ville, 2017]; in finance, causal dependencies among assets shift during market regime changes [Papana et al., 2017]; and in cardiac fibrillation, preferential conduction pathways may underpin the fibrillation mechanism [Handa, X. Li, et al., 2020].

In the case of cardiac fibrillation, the electrophenotype spectrum framework (Sec. 3.3) posits that each patient’s fibrillation is maintained by patient-specific propagation patterns whose complexity ranges from organised, focal drivers to fully disorganised multi-wavelet dynamics [F. S. Ng et al., 2020]. Characterising these patterns requires identifying the preferential pathways of electrical propagation (the functional connectivity) and assessing whether these pathways are sufficiently stable to serve as viable ablation targets [Handa, X. Li, et al., 2020]. Since AF dynamics exhibit substantial temporal variability, with transitions between organised and disorganised states [Everett IV et al., 2001; Ravelli et al., 2007] and non-stationary characteristics [Stridh et al., 2001; M. Sun et al., 2020], batch graph learning approaches that estimate a single static topology from an entire dataset may not adequately capture the temporal evolution of these patterns.

Batch approaches to graph topology estimation include sparse inverse covariance methods [Friedman, Hastie, and Tibshirani, 2007], signal-smoothness formulations [Xiaowen Dong, Thanou, et al., 2016], and causal graph learning methods that test whether the history of one signal improves prediction of another [Granger, 1969]. The latter framework extends naturally to graph-structured data through the time-vertex autoregressive system identification approach of Mei and Moura (2016), where the causal topology is parameterised as the graph shift operator within a CGP model (Sec. 2.3.1). Deep learning methods have also been developed to learn the graph structure end-to-end [Cini, Zambon, and Alippi, 2023; Shang and J. Chen, 2021], including the graph learning module within COREL (Sec. 6.3.2), but their computational demands and reliance on large training datasets preclude real-time online estimation of this topology from streaming data.

When the topology is itself non-stationary, online algorithms that track changes with constant per-iteration cost become necessary. Adaptive signal processing techniques (Sec. 2.1.4) provide a

natural foundation: algorithms such as LMS [Widrow and Hoff, 1988], RLS [Sayed, 2003], and RLS-like LMS variants [Varidhissai and Danilo P Mandic, 2017] provide recursive update rules for tracking time-varying parameters from streaming data. Recent works have introduced online schemes for estimating causal graphs from sparse VAR processes, such as TISO and TIRSO [Zaman, Ramos, et al., 2020], and sparse structural equation model processes [Baingana and Giannakis, 2016]. However, achieving true sparsity in online settings remains a fundamental challenge: existing methods struggle to explicitly zero out non-causal elements despite employing sparsity-promoting regularisation [Taheri and Vorobyov, 2011]. This is a critical limitation for topology estimation, since the binary presence or absence of edges is precisely what determines graph structure and its causal interpretation. While adaptive filtering techniques have been extended to GSP [Di Lorenzo, Barbarossa, et al., 2016; Yan, Kuruoglu, and Altinkaya, 2022; Di Lorenzo, Banelli, et al., 2017], this work has focused primarily on signal estimation and sampling on graphs rather than topology learning.

This chapter introduces ADACGP, which bridges the gap between the offline CGP framework of Sec. 2.3.1 and the adaptive filtering principles of Sec. 2.1.4 to enable online graph topology estimation from streaming data. The contributions of this chapter are fourfold: (i) an online time-vertex adaptive filtering algorithm that extends the batch CGP identification of Mei and Moura (2016) to streaming data, with a focus on accurate topology estimation; (ii) a variable splitting approach for achieving true sparsity in online graph shift operator estimation, enabling reliable identification of non-zero causal elements and efficient computation through sparse matrix operations; (iii) a rigorous empirical analysis on synthetic data demonstrating convergence and superiority over adaptive VAR baselines, with prediction error optimisation enabling reliable hyperparameter selection and sparsity estimation; and (iv) an application to VF recordings across increasing anti-arrhythmic drug concentrations, demonstrating the ability of ADACGP to capture the structure and stability of cardiac fibrillation.

7.2 Methodology

7.2.1 Problem Formulation

Consider a graph, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, with N nodes and graph shift operator, $\mathbf{S} \in \mathbb{R}^{N \times N}$ (Sec. 2.2.1). A time-varying graph signal, $\mathbf{x}[t] \in \mathbb{R}^N$, associates a scalar value with each node at time step t . The CGP model (Sec. 2.3.1) expresses the signal as a sum of graph-filtered lagged observations,

$$\mathbf{x}[t] = \sum_{p=1}^P \Psi_p(\mathbf{S}) \mathbf{x}[t-p] + \mathbf{e}[t], \quad (7.1)$$

where $\Psi_p(\mathbf{S}) = \sum_{k=0}^p h_{p,k} \mathbf{S}^k$ is a polynomial graph filter of order at most p (Sec. 2.2.3), $h_{p,k}$ are scalar filter coefficients, and $\mathbf{e}[t] \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_N)$ is innovation noise. The coupling of spatial reach to temporal lag, whereby the graph filter at lag p has order at most p , encodes a finite propagation speed of information through the graph: a signal emitted p time steps ago can influence at most nodes within a p -hop neighbourhood (Fig. 2.8). As polynomial graph filters, the $\Psi_p(\mathbf{S})$ are shift-invariant, commuting with both $\mathbf{S}$ and with each other, $[\Psi_i(\mathbf{S}), \Psi_j(\mathbf{S})] = \mathbf{0}$ for all i, j (Sec. 2.2.3).

The graph topology estimation problem is to recover $\mathbf{S}$ from streaming observations, $\{\mathbf{x}[t]\}_{t=1}^T$, under the assumption that the data are generated by a process of the form (7.1). Unlike the signal reconstruction and uncertainty estimation tasks of Chapters 5 and 6, the focus here is strictly on system identification: estimating $\mathbf{S}$ and the remaining parameters of (7.1) from the data.

Remark 14. *The CGP model in (7.1) is inherently weakly stationary due to its formulation as a sum of shift-invariant graph filters [Marques, Segarra, et al., 2017] (Sec. 2.3.1). While stationarity assumptions enable the definition of meaningful optimality criteria for filter design, real-world scenarios often involve non-stationary signals and time-varying graph topologies. This motivates the adaptive approach developed in this chapter, which tracks gradual changes through recursive optimisation using the forgetting factor mechanism of Sec. 2.1.4.*

7.2.2 Objectives for Online Optimisation

The CGP model in (7.1) can be formulated as a multivariate linear regression problem for estimating $\mathbf{S}$ and $\{\mathbf{h}_p\}_{p=1}^P$. The RLS-like LMS scheme [Varidhissai and Danilo P Mandic, 2017] is adopted for recursive optimisation, which incorporates a forgetting factor λ into the LMS objective and achieves fast convergence and tracking performance similar to RLS while being more numerically stable, as it does not involve matrix inversion (Sec. 2.1.4). The least squares problem is formulated as

$$\min_{\mathbf{S}, \mathbf{h}} \frac{1}{2} \sum_{\tau=1}^t \lambda^{t-\tau} \left\| \mathbf{x}[\tau] - \sum_{p=1}^P \Psi_p(\mathbf{S}) \mathbf{x}[\tau-p] \right\|_2^2, \quad (7.2)$$

where $\mathbf{h} = [\mathbf{h}_1^\top, \dots, \mathbf{h}_P^\top]^\top \in \mathbb{R}^M$ with $M = P(P+3)/2$, $\mathbf{x}[\tau] = \mathbf{0}$ for $\tau \leq 0$, and $\lambda \in (0, 1]$. This represents a non-convex optimisation problem due to the polynomial in $\mathbf{S}$. Following Mei and Moura (2016), the problem is cast into alternating steps of convex regularised least squares sub-problems.

Solving for $\Psi_p = H_p(\mathbf{S}, \mathbf{h}_p)$

The objective is first reformulated to solve for Ψ_p instead of $\mathbf{S}$ and $\mathbf{h}$ directly. This transformation makes the problem quadratic in Ψ_p . Incorporating sparsity regularisation yields

$$\min_{\Psi} \frac{1}{2} \sum_{\tau=1}^t \lambda^{t-\tau} \left\| \mathbf{x}[\tau] - \sum_{p=1}^P \Psi_p \mathbf{x}[\tau-p] \right\|_2^2 + \sum_{p=1}^P \mu_p \|\text{vec}(\Psi_p)\|_1, \quad (7.3)$$

where $\Psi = [\Psi_1, \dots, \Psi_P] \in \mathbb{R}^{N \times NP}$ is a concatenation of the P graph filters, $\text{vec}(\cdot)$ stacks the columns of the matrix it acts on, $\|\cdot\|_1$ is the ℓ_1 norm, and μ_p controls the degree of sparsity of the corresponding $\hat{\Psi}_p$.

To enforce the shift-invariant property of polynomial graph filters (Sec. 2.2.3), a soft constraint is added to arrive at

$$\min_{\Psi} \frac{1}{2} \sum_{\tau=1}^t \lambda^{t-\tau} \left\| \mathbf{x}[\tau] - \sum_{p=1}^P \Psi_p \mathbf{x}[\tau-p] \right\|_2^2 + \sum_{p=1}^P \mu_p \|\text{vec}(\Psi_p)\|_1 + \gamma \sum_{i \neq j} \|\Psi_i - \Psi_j\|_F^2, \quad (7.4)$$

where the additional term, weighted by γ , enforces each pair of graph filters to commute. This formulation results in a quartic programming problem. However, as noted by Mei and Moura (2016), the objective is multi-convex and naturally leads to block coordinate descent: when all Ψ_j except Ψ_i are held constant, the problem becomes convex in Ψ_i .

Estimating $\mathbf{S}$ from $\hat{\Psi}_1$

Given that an order-1 graph filter, $\Psi_1 = h_{1,0}\mathbf{I} + h_{1,1}\mathbf{S}$, is a linear function of $\mathbf{S}$, its estimate, $\hat{\Psi}_1$, provides a reasonable yet biased approximation of $\mathbf{S}$. To refine this estimate, the following regularised least squares sub-problem is solved

$$\min_{\mathbf{S}} \frac{1}{2} \left\| \hat{\Psi}_1 - \mathbf{S} \right\|_2^2 + \mu_1 \|\text{vec}(\mathbf{S})\|_1 + \gamma \sum_{p=2}^P \|\mathbf{S} - \hat{\Psi}_p\|_F^2, \quad (7.5)$$

where the final term enforces the shift-invariant property by ensuring that $\hat{\mathbf{S}}$ commutes with all $\hat{\Psi}_p$.

When (7.4) is used to calculate $\hat{\Psi}_1$, the sub-problem in (7.5) may be bypassed by the approximation $\hat{\mathbf{S}} = \hat{\Psi}_1$, since shift-invariance has already been enforced. This variant is denoted as **Path 2**, detailed in subsequent sections.

Estimating $\mathbf{h}$

Upon estimating $\mathbf{S}$, the initial objective in (7.2) can be reframed as a quadratic optimisation with respect to $\mathbf{h}$. Substituting the polynomial graph filter form into (7.2) and assuming sparsity in $\mathbf{h}$ gives

$$\min_{\mathbf{h}} \frac{1}{2} \sum_{\tau=1}^t \lambda^{t-\tau} \|\mathbf{x}[\tau] - \mathbf{Y}[\tau] \mathbf{h}\|_2^2 + \eta \|\mathbf{h}\|_1, \quad (7.6)$$

where η controls the degree of sparsity and

$$\mathbf{Y}[t] = [\mathbf{x}[t-1], \hat{\mathbf{S}}\mathbf{x}[t-1], \dots, \mathbf{x}[t-P], \dots, \hat{\mathbf{S}}^P\mathbf{x}[t-P]]. \quad (7.7)$$

The dimensions of $\mathbf{Y}[t]$ are $N \times M$, where $M = P(P+3)/2$. Despite the apparent size of M , practical applications often involve a low model order satisfying $M < N$. This step, while informative for model identification, is not critical to the primary goal of recovering $\mathbf{S}$.

7.2.3 Online Estimation Algorithm

The ADACGP algorithm is now introduced, designed to achieve sparse online solutions to the objective in (7.2). While existing online methods such as ℓ_1 -regularised LMS [Yilun Chen, Y. Gu, and Alfred O Hero, 2009] and oracle shrinkage algorithms [Lamare and Sampaio-Neto, 2014] achieve low MSE, they rarely produce truly sparse solutions compared to offline methods like basis pursuit [Taheri and Vorobyov, 2011]. This is a critical limitation for estimating $\mathbf{S}$, since sparsity directly determines the presence or absence of edges in the graph topology, driving the causal interpretation.

To overcome this, the offline variable splitting approach of Schmidt, Fung, and Rosales (2007) is adapted to the online objective in (7.2), reformulating the alternating ℓ_1 -regularised sub-problems by splitting the target variables ($\hat{\Psi}$ and $\hat{\mathbf{S}}$) into their positive and negative components as

$$\hat{\Psi} \triangleq \hat{\Psi}_+ - \hat{\Psi}_-, \quad (7.8)$$

$$\hat{\mathbf{S}} \triangleq \hat{\mathbf{S}}_+ - \hat{\mathbf{S}}_-, \quad (7.9)$$

where $(\cdot)_+ \geq \mathbf{0}$ and $(\cdot)_- \geq \mathbf{0}$ contain only the positive and negative parts of $(\cdot)$, respectively. The ℓ_1 -norm can then be expressed as

$$\|\hat{\Psi}\|_1 = \text{Tr}(\mathbf{1}_{N \times N} \hat{\Psi}_+) + \text{Tr}(\mathbf{1}_{N \times N} \hat{\Psi}_-),$$

and similarly for $\hat{\mathbf{S}}$, where $\text{Tr}(\cdot)$ is the trace operator and $\mathbf{1}_{N \times N}$ is a matrix of unities. These refor-

mulations convert the first two sub-problems into non-negativity constrained optimisations [Schmidt, Fung, and Rosales, 2007], solved via projected stochastic gradient descent using the recursive update formulae derived next.

Remark 15. *The variable splitting approach in (7.8) and (7.9) offers potential for incorporating prior knowledge about the graph shift operator structure. For example, when the graph shift operator represents an adjacency matrix (elements in $\{0, 1\}$), setting $\hat{\Psi}_- = \hat{S}_- = \mathbf{0}$ naturally enforces this constraint as an inductive bias. While Laplacian matrices could be split into their positive diagonal and negative off-diagonal elements, their zero row sum proves challenging to maintain iteratively without resorting to Lagrangian methods. To maintain generality, the unconstrained case is studied following Mei and Moura (2016).*

Updating $\Psi[t]$

To minimise (7.4) at time t , the gradient with respect to the positive part of each p -th block of the graph filter, $(\hat{\Psi}_p)_+$, is given by

$$\nabla_{(\hat{\Psi}_p)_+}^{(t)} = \sum_{\tau=1}^t \lambda^{t-\tau} \left(\sum_{k=1}^P \hat{\Psi}_k[t-1] \mathbf{x}[\tau-k] \mathbf{x}[\tau-p]^\top - \mathbf{x}[\tau] \mathbf{x}[\tau-p]^\top \right) + \mu_p[t] \mathbf{1}_{N \times N} + \gamma \mathbf{Q}_p[t], \quad (7.10)$$

where the gradient of the commutativity term is

$$\mathbf{Q}_p[t+1] = \sum_{k \neq p}^P \left([\hat{\Psi}_p[t], \hat{\Psi}_k[t]] \hat{\Psi}_k[t]^\top - \hat{\Psi}_k[t]^\top [\hat{\Psi}_p[t], \hat{\Psi}_k[t]] \right). \quad (7.11)$$

The quantities $\hat{\Psi}[t]$, $\mathbf{M}[t]$, and $\mathbf{Q}[t]$ are defined as the concatenations of the graph filters, sparsity terms, and commutativity terms over all P , respectively, given by

$$\hat{\Psi}[t] \triangleq [\hat{\Psi}_1[t], \hat{\Psi}_2[t], \dots, \hat{\Psi}_P[t]] := \hat{\Psi}_+[t] - \hat{\Psi}_-[t],$$

$$\mathbf{M}[t] \triangleq [\mu_1[t] \mathbf{1}_{N \times N}, \mu_2[t] \mathbf{1}_{N \times N}, \dots, \mu_P[t] \mathbf{1}_{N \times N}],$$

and

$$\mathbf{Q}[t] \triangleq [\mathbf{Q}_1[t], \mathbf{Q}_2[t], \dots, \mathbf{Q}_P[t]],$$

where $\hat{\Psi}[t]$ is expressed as the difference between its positive and negative parts, as in (7.8), and $\hat{\Psi}[t]$, $\mathbf{M}[t]$ and $\mathbf{Q}[t]$ are matrices in $\mathbb{R}^{N \times NP}$.

Next, $\mathbf{R}[t] \in \mathbb{R}^{NP \times NP}$ denotes the accumulated correlation matrix of the lagged input signals, and $\mathbf{P}[t] \in \mathbb{R}^{N \times NP}$ denotes the accumulated cross-correlation matrix between the input signals $\mathbf{x}_P[\tau]$ and the desired responses $\mathbf{x}[t]$, where the lagged signal vector is given by $\mathbf{x}_P[\tau] \triangleq [\mathbf{x}[\tau-1]^\top, \mathbf{x}[\tau-2]^\top, \dots, \mathbf{x}[\tau-P+1]^\top]^\top$,

$\dots, \mathbf{x}[\tau-P]^\top]^\top \in \mathbb{R}^{NP}$.

The recursive update formulae for $\mathbf{R}[t]$ and $\mathbf{P}[t]$ are then given by

$$\mathbf{R}[t] \triangleq \sum_{\tau=1}^t \lambda^{t-\tau} \mathbf{x}_P[\tau] \mathbf{x}_P[\tau]^\top = \lambda \mathbf{R}[t-1] + \mathbf{x}_P[t] \mathbf{x}_P[t]^\top,$$

and

$$\mathbf{P}[t] \triangleq \sum_{\tau=1}^t \lambda^{t-\tau} \mathbf{x}[\tau] \mathbf{x}_P[\tau]^\top = \lambda \mathbf{P}[t-1] + \mathbf{x}[t] \mathbf{x}_P[t]^\top.$$

These expressions are then substituted into (7.10) to give

$$\mathbf{G}[t] \triangleq \left[\nabla_{(\hat{\Psi}_1)}^{(t)}, \dots, \nabla_{(\hat{\Psi}_P)}^{(t)} \right] \triangleq \hat{\Psi}[t-1] \mathbf{R}[t] - (\mathbf{P}[t] - \gamma \mathbf{Q}[t]), \quad (7.12)$$

where it can be shown straightforwardly that $\nabla_{(\hat{\Psi}_p)_+}^{(t)} = \nabla_{(\hat{\Psi}_p)}^{(t)}$ and $\nabla_{(\hat{\Psi}_p)_-}^{(t)} = -\nabla_{(\hat{\Psi}_p)}^{(t)}$ for all p .

Finally, the updates to the first sub-problem can be expressed as a gradient projection, that is

$$\hat{\Psi}_+[t] = \left(\hat{\Psi}_+[t-1] - (\mathbf{M}[t] + \mathbf{G}[t])(\mathbf{A}[t] \otimes \mathbf{I}_{N \times N}) \right)_+,$$

where $\mathbf{A}[t] = \text{diag}(\alpha_1, \alpha_2, \dots, \alpha_P) \in \mathbb{R}^{P \times P}$ is a diagonal matrix of step-sizes for each filter block.

Similarly, the update equation for $\hat{\Psi}_-[t]$ can be obtained as

$$\hat{\Psi}_-[t] = \left(\hat{\Psi}_-[t-1] - (\mathbf{M}[t] - \mathbf{G}[t])(\mathbf{A}[t] \otimes \mathbf{I}_{N \times N}) \right)_+.$$

Then, finally $\hat{\Psi}[t] = \hat{\Psi}_+[t] - \hat{\Psi}_-[t]$ can be set.

Having derived the update equations for $\hat{\Psi}[t]$, the computational complexity per iteration is now analysed. The cost is dominated by the gradient calculation in (7.12). Computing $\hat{\Psi}[t-1] \mathbf{R}[t]$ and $\mathbf{Q}[t]$ requires $\mathcal{O}(N^3 P^2)$ operations in the dense case. Since the algorithm explicitly zeros out elements of $\hat{\Psi}[t]$, sparse matrix operations could be used. With $s[t] = \|\hat{\Psi}[t]\|_0$ total non-zeros and $s_p[t] = \max_p \|\hat{\Psi}_p[t]\|_0$ maximum non-zeros per block at time t , the complexity reduces to $\mathcal{O}(s[t-1]NP)$ for $\hat{\Psi}[t-1] \mathbf{R}[t]$ and $\mathcal{O}(s_p[t-1]^2 P)$ for $\mathbf{Q}[t]$. Since $s[t-1] \leq N^2 P$, $s_p[t-1] \leq N^2$, and $s[t-1] \geq s_p[t-1]$, the total complexity simplifies to $\mathcal{O}(s[t-1]NP)$, which scales more favourably for large data applications.

Updating $\mathbf{S}[t]$

The graph shift operator, $\mathbf{S}$, is estimated from $\hat{\Psi}[t]$ through one of two paths. **Path 2** provides a simple but biased estimate by setting $\hat{\mathbf{S}}[t] = \hat{\Psi}_1[t]$. **Path 1**, detailed below, optimises (7.5) to

obtain direct estimates of $\mathbf{S}$ with reduced bias.

In Path 1, (7.12) in the first sub-problem is simplified by setting $\mathbf{Q}[t] = \mathbf{0}$ for all time steps t . This makes the first sub-problem convex by deferring the enforcement of shift-invariance to this phase instead. To optimise (7.5), the variable splitting approach from (7.9) is followed to express $\hat{\mathbf{S}}[t]$ as the difference between its positive and negative parts. Computing the gradient of (7.5) with respect to the positive elements of $\hat{\mathbf{S}}[t]$ gives

$$\mathbf{V}[t] = \hat{\mathbf{S}}[t-1] - (\hat{\Psi}_1[t] - \gamma \mathbf{C}[t]), \quad (7.13)$$

where the gradient of the commutative term is given by

$$\mathbf{C}[t] = \sum_{k=2}^P \left(\left[\hat{\mathbf{S}}[t-1], \hat{\Psi}_k[t] \right] \hat{\Psi}_k[t]^\top - \hat{\Psi}_k[t]^\top \left[\hat{\mathbf{S}}[t-1], \hat{\Psi}_k[t] \right] \right). \quad (7.14)$$

The parameter updates can again be expressed as gradient projections

$$\hat{\mathbf{S}}_+[t] = (\hat{\mathbf{S}}_+[t-1] - \beta[t](\mu_1[t]\mathbf{1}_{N \times N} + \mathbf{V}[t]))_+,$$

and

$$\hat{\mathbf{S}}_-[t] = (\hat{\mathbf{S}}_-[t-1] - \beta[t](\mu_1[t]\mathbf{1}_{N \times N} - \mathbf{V}[t]))_+,$$

where $\beta[t]$ is the step-size, and $\mu_1[t]$ is the sparsity parameter from the first sub-problem. Then, $\hat{\mathbf{S}}[t] = \hat{\mathbf{S}}_+[t] - \hat{\mathbf{S}}_-[t]$.

The computational complexity for updating $\hat{\mathbf{S}}[t]$ varies between paths. While Path 2 requires $\mathcal{O}(N^2)$ operations for direct assignment of $\hat{\mathbf{S}}[t] = \hat{\Psi}_1[t]$, Path 1 has higher complexity of $\mathcal{O}(N^3P)$ due to the computation of the commutative term $\mathbf{C}[t]$ in the dense case. However, when matrices $\hat{\mathbf{S}}$ and $\hat{\Psi}_k$ are sparse in (7.14), Path 1 complexity can be reduced to $\mathcal{O}(s_p[t]^2P)$, where $s_p[t]$ is as defined previously.

Algorithm 1 summarises the derived adaptive algorithm for learning $\hat{\mathbf{S}}[t]$. Path 1 ignores Step 9 by setting $\mathbf{Q}[t] = \mathbf{0}$, executing only Step 18 to account for shift-invariance. Conversely, Path 2 includes Step 9 but skips Steps 18-22, directly assigning $\hat{\mathbf{S}}[t] = \hat{\Psi}_1[t]$ instead. These two paths will be tested experimentally in subsequent sections.

Algorithm 1: Identifying the topology of $\hat{\mathbf{S}}$ ($\hat{\mathbf{S}}^*$)

Input : $\mathbf{x}, P$

Output: $\hat{\Psi}, \hat{\mathbf{S}}^*$

```

1 Initialise  $\hat{\Psi}[0] = \hat{\Psi}_+[0] = \hat{\Psi}_-[0] = \mathbf{P}[0] = \mathbf{Q}[1] = \mathbf{0}$ ,  $\hat{\mathbf{S}}[0] = \hat{\mathbf{S}}_+[0] = \hat{\mathbf{S}}_-[0] = \mathbf{0}$  and  $\mathbf{R}[0] = \mathbf{0}$ ;
2  $t = 0$ ;
3 do
4    $t = t + 1$ ;

5   Solving for  $\hat{\Psi}[t]$ ;
6    $\mathbf{x}_P[t] = [\mathbf{x}[t-1]^\top, \mathbf{x}[t-2]^\top, \dots, \mathbf{x}[t-P]^\top]^\top$ ;
7    $\mathbf{R}[t] = \lambda \mathbf{R}[t-1] + \mathbf{x}_P[t] \mathbf{x}_P[t]^\top$ ;
8    $\mathbf{P}[t] = \lambda \mathbf{P}[t-1] + \mathbf{x}[t] \mathbf{x}_P[t]^\top$ ;
9    $\mathbf{Q}[t] = [\mathbf{Q}_1[t], \mathbf{Q}_2[t], \dots, \mathbf{Q}_P[t]]$  with  $\mathbf{Q}_p[t]$  according to (7.11);
10  calculate  $\boldsymbol{\mu}[t]$ ;
11   $\mathbf{M}[t] = [\mu_1[t] \mathbf{1}_{N \times N}, \mu_2[t] \mathbf{1}_{N \times N}, \dots, \mu_P[t] \mathbf{1}_{N \times N}]$ ;
12   $\mathbf{G}[t] = \hat{\Psi}[t-1] \mathbf{R}[t] - (\mathbf{P}[t] - \gamma \mathbf{Q}[t])$ ;
13  calculate  $\mathbf{A}[t]$ ;
14   $\hat{\Psi}_+[t] = \left( \hat{\Psi}_+[t-1] - (\mathbf{M}[t] + \mathbf{G}[t])(\mathbf{A}[t] \otimes \mathbf{I}_{N \times N}) \right)_+$ ;
15   $\hat{\Psi}_-[t] = \left( \hat{\Psi}_-[t-1] - (\mathbf{M}[t] - \mathbf{G}[t])(\mathbf{A}[t] \otimes \mathbf{I}_{N \times N}) \right)_+$ ;
16   $\hat{\Psi}[t] = \hat{\Psi}_+[t] - \hat{\Psi}_-[t]$ ;

17  Estimating  $\hat{\mathbf{S}}[t]$ ;
18   $\mathbf{C}[t]$  according to (7.14);
19   $\mathbf{V}[t] = \hat{\mathbf{S}}[t-1] - (\hat{\Psi}_1[t] - \gamma \mathbf{C}[t])$ ;
20   $\hat{\mathbf{S}}_+[t] = \left( \hat{\mathbf{S}}_+[t-1] - \beta[t](\mu_1[t] \mathbf{1}_{N \times N} + \mathbf{V}[t]) \right)_+$ ;
21   $\hat{\mathbf{S}}_-[t] = \left( \hat{\mathbf{S}}_-[t-1] - \beta[t](\mu_1[t] \mathbf{1}_{N \times N} - \mathbf{V}[t]) \right)_+$ ;
22   $\hat{\mathbf{S}}[t] = \hat{\mathbf{S}}_+[t] - \hat{\mathbf{S}}_-[t]$ ;

23 while  $t < T^*$  (an epoch with steady-state reached);
24  $\hat{\mathbf{S}}^* = \hat{\mathbf{S}}[T^*]$ .

```

Estimating $\mathbf{h}[t]$

To fully identify the CGP model in (7.1), the graph filter coefficients $\hat{\mathbf{h}}[t]$ remain to be estimated. This phase is optional since the primary objective of this chapter is to estimate the structure of the graph shift operator. Notice that, $\hat{\mathbf{S}}[t]$ must be *debiased* before estimating $\hat{\mathbf{h}}[t]$, due to the heavy regularisation involved. Debiasing is performed by fixing and optimising only the non-zero elements of $\hat{\mathbf{S}}[t]$ using least squares. Note that reducing the sample size in this way can risk distorting the noise distribution from normality, potentially compromising the minimum MSE criterion for noisy or small datasets [Donoho, 1995].

After debiasing $\hat{\mathbf{S}}[t]$, the filter coefficients $\hat{\mathbf{h}}[t]$ are estimated. Unlike zeros in the graph shift operator, which indicate absence of edges and thus have causal significance, sparsity in $\hat{\mathbf{h}}$ serves mainly to simplify the model structure. Given the more relaxed sparsity requirements, GAR-LMS [Taheri and Vorobyov, 2011] is employed to arrive at the recursive update equation of $\hat{\mathbf{h}}$, given by

$$\hat{\mathbf{h}}[t] = \hat{\mathbf{h}}[t-1] + \rho[t] (\mathbf{D}[t]\hat{\mathbf{h}}[t-1] - \mathbf{u}[t] + \eta[t]\mathbf{b}[t]), \quad (7.15)$$

where

$$\mathbf{D}[t] = \lambda \mathbf{D}[t-1] + \mathbf{Y}[t]^\top \mathbf{Y}[t],$$

$$\mathbf{u}[t] = \lambda \mathbf{u}[t-1] + \mathbf{Y}[t]^\top \mathbf{x}[t],$$

$$\mathbf{b}[t] : b_i[t] = \frac{\text{sign}(\hat{h}_i[t-1])}{\varepsilon + \hat{h}_i[t-1]},$$

where $\mathbf{Y}[t]$ contains the lagged graph shifted signals as in (7.7), $\mathbf{D}[t] \in \mathbb{R}^{M \times M}$ tracks correlations between the shifted signals, $\mathbf{u}[t] \in \mathbb{R}^M$ accumulates cross-correlations with the desired response, $\mathbf{b}[t]$ is the sparsity-promoting re-weighting vector, $\rho[t]$ is the step-size, and ε prevents division by zero. This step could be further simplified by only taking the instantaneous samples into (7.15), that is, $\lambda = 0$, to yield

$$\hat{\mathbf{h}}[t] = \hat{\mathbf{h}}[t-1] + \rho[t] (\mathbf{Y}[t]^\top \mathbf{e}[t] + \eta[t]\mathbf{b}[t])$$

where

$$\mathbf{e}[t] = \mathbf{x}[t] - \mathbf{Y}[t]\hat{\mathbf{h}}[t-1].$$

Algorithm 2 summarises the derived adaptive algorithm for debiasing $\hat{\mathbf{S}}[t]$ and estimating $\hat{\mathbf{h}}[t]$. The cost of the debiasing phase is dominated by the term $\hat{\Psi}[t-1]\mathbf{R}[t]$, which has complexity $\mathcal{O}(N^3P^2)$ in the dense case or $\mathcal{O}(s[t-1]NP)$ with sparse operations. The cost of updating $\hat{\mathbf{h}}[t]$ is dominated by computing $\mathbf{Y}[t]^\top \mathbf{Y}[t]$, requiring $\mathcal{O}(NP^4)$ operations for $\mathbf{Y}[t] \in \mathbb{R}^{N \times M}$ with $M = P(P+3)/2$.

While appearing formidable, P is typically much smaller than N , so this update is unlikely to be the computational bottleneck in the overall algorithm.

Algorithm 2: Determining the *unbiased* $\hat{\mathbf{S}}$ and $\hat{\mathbf{h}}$

Input : $\mathbf{x}, P, \delta$
Output: $\hat{\mathbf{S}}, \hat{\mathbf{h}}$

```

1 All recursive variables resume from Algorithm 1;
2  $t = T^*$ ;
3 do
4    $t = t + 1$ ;
5   Debiasing  $\hat{\mathbf{S}}[t]$ ;
6    $\mathbf{R}[t] = \lambda \mathbf{R}[t-1] + \mathbf{x}_P[t] \mathbf{x}_P[t]^\top$ ;
7    $\mathbf{P}[t] = \lambda \mathbf{P}[t-1] + \mathbf{x}[t] \mathbf{x}_P[t]^\top$ ;
8    $\mathbf{G}[t] = (\hat{\Psi}[t-1] \mathbf{R}[t] - \mathbf{P}[t])_{\hat{\mathbf{S}}}$  where  $(\cdot)_{\hat{\mathbf{S}}}$  is the projection to non-zero elements of  $\hat{\Psi}$  considering  $\hat{\mathbf{S}}$ ;
9   calculate  $\mathbf{A}[t]$ ;
10   $\hat{\Psi}[t] = \hat{\Psi}[t-1] - \mathbf{G}[t](\mathbf{A}[t] \otimes \mathbf{I}_{N \times N})$ ;
11  Setting  $\hat{\mathbf{S}}[t] = \hat{\Psi}_1[t]$ ;
12  Estimating  $\hat{\mathbf{h}}[t]$ ;
13   $\mathbf{Y}[t] = [\mathbf{x}[t-1], \hat{\mathbf{S}}[t] \mathbf{x}[t-1], \dots, \mathbf{x}[t-P], \dots, \hat{\mathbf{S}}[t]^P \mathbf{x}[t-P]]$ ;
14   $\mathbf{e}[t] = \mathbf{x}[t] - \mathbf{Y}[t] \hat{\mathbf{h}}[t-1]$ ;
15   $\mathbf{b}[t] : b_i[t] = \frac{\text{sign}(\hat{h}_i[t])}{\varepsilon + \hat{h}_i[t]}$ ;
16   $\hat{\mathbf{h}}[t] = \hat{\mathbf{h}}[t-1] + \rho[t] (\mathbf{Y}[t]^\top \mathbf{e}[t] + \eta[t] \mathbf{b}[t])$ ;
17 while  $t < T$  (a terminal epoch);
18  $\hat{\mathbf{S}} = \hat{\mathbf{S}}[T], \hat{\mathbf{h}} = \hat{\mathbf{h}}[T]$ 

```

Tuning Hyperparameters

The algorithm requires tuning several hyperparameters: regularisation constants $\boldsymbol{\mu}[t] = [\mu_1[t], \dots, \mu_P[t]]^\top$, $\eta[t]$, γ , and ε ; step-sizes $\mathbf{A}[t]$, $\beta[t]$, and $\rho[t]$; and the forgetting factor λ .

Prior knowledge guides many of these selections. Following S.-J. Kim, Koh, et al. (2007), the ℓ_1 -norm related constants are set as $\mu_p[t] = \mu_p \|\mathbf{P}_p[t] - \gamma \mathbf{Q}_p[t]\|_\infty$ and $\eta[t] = \eta \|\mathbf{Y}[t]^\top \mathbf{x}[t]\|_\infty$, where $\mathbf{P}_p[t]$ is the p -th block of $\mathbf{P}[t]$. Step-sizes can be determined using Armijo line search [Boukis, Danilo P Mandic, et al., 2010] or adaptive signal processing theory for stable convergence. The entries of $\boldsymbol{\mu}$ could decrease with p to reflect the decreasing sparsity of higher-order graph filters (Sec. 2.2.3). The forgetting factor λ should be close to 1 to balance adaptability and stability, consistent with the principles established in Sec. 2.1.4. However, parameters η and γ are rather unconstrained and require empirical tuning.

Discussion on Convergence

A rigorous convergence analysis for batch CGP identification using the same alternating steps of convex regularised least squares sub-problems was presented by Mei and Moura (2016). Their analysis bounded the average excess prediction risk, defined as the difference between prediction

error using the estimated and true parameters. This error scales as $\mathcal{O}(\sqrt{\log N/K})$ for estimating the graph shift operator, $\mathbf{S}$, and as $\mathcal{O}(\sqrt{\log N/K^{2(\nu-\beta)}})$ for estimating the filter coefficients, $\mathbf{h}$, where N is the network size, K is the number of time observations, and ν, β are parameters satisfying $0 < \beta < \nu < 1/2$.

ADACGP extends this framework through variable splitting as expressed in (7.8) and (7.9), while maintaining the same core optimisation structure. The variable splitting reformulates the ℓ_1 -regularised problem as a non-negativity constrained optimisation [Schmidt, Fung, and Rosales, 2007], aligning with two-metric projection methods [Gafni and Bertsekas, 1984]. For such methods, Gafni and Bertsekas (1984) established that for non-critical points there exists a step-size that guarantees descent, and that every limit point of the generated sequence is a critical point. When applied to the non-negativity constraints resulting from variable splitting, these results provide theoretical support for the algorithm’s convergence properties, with projected gradient theory suggesting similar convergence behaviour under standard convexity and step size conditions. The effectiveness of this technique for ℓ_1 -regularised problems has been empirically demonstrated by [Schmidt, Fung, and Rosales, 2007], showing consistent improvements in convergence speed and computational efficiency.

The relationship between batch and online optimisation is well-established. For adaptive filtering, LMS converges to the Wiener–Hopf solution under small step-sizes and stationarity [Widrow and Hoff, 1988], and RLS variants achieve faster convergence rates [Sayed, 2003; Varidhissai and Danilo P Mandic, 2017]. This suggests the convergence properties established for the batch CGP algorithm [Mei and Moura, 2016] and variable splitting [Gafni and Bertsekas, 1984] should extend to the adaptive setting, given stationarity, appropriate forgetting factor, and sufficient iterations. Despite these theoretical guarantees, practical limitations remain: ℓ_1 -regularisation introduces shrinkage bias [S.-J. Kim, Koh, et al., 2007], and the commutativity term in (7.4) adds further complexity. The analysis is therefore complemented with empirical convergence metrics, evaluating ADACGP’s ability to correctly identify the graph topology through precision, recall, and F1 scores for non-zero elements in $\mathbf{S}$.

7.3 Experiments

The experimental evaluation proceeds in two stages. First, ADACGP is validated in a controlled synthetic environment across four graph topologies, assessing convergence behaviour, topology recovery accuracy, and computational scaling relative to adaptive and batch baselines. Second, the algorithm is applied to *ex vivo* optical mapping recordings of rat ventricular fibrillation, demon-

strating its ability to track dynamic changes in functional connectivity and characterise the structure and stability of cardiac fibrillation under varying anti-arrhythmic drug concentrations.

7.3.1 Controlled Synthetic Environment

To evaluate ADACGP in a setting where ground-truth graph topology is known, experiments are conducted on synthetically generated CGP data across four representative graph topologies of varying structure and sparsity. This controlled environment allows rigorous assessment of topology recovery accuracy and convergence behaviour, and enables principled comparison against adaptive and batch baselines. Hyperparameters are selected by minimising prediction error on the observed signals, without access to the ground-truth topology, reflecting the conditions under which the algorithm would be deployed in practice.

Evaluation metrics. The performance and convergence characteristics of ADACGP across different graph topologies are assessed using three primary metrics: the prediction error from the estimated graph filter,

$$\text{NMSE}(\mathbf{x}_\Psi) = \frac{\|\mathbf{x}[t] - \hat{\Psi}[t-1] \mathbf{x}_P[t]\|_2^2}{\|\mathbf{x}[t]\|_2^2}, \quad (7.16)$$

the prediction error from the estimated graph filter coefficients,

$$\text{NMSE}(\mathbf{x}_h) = \frac{\|\mathbf{x}[t] - \mathbf{Y}[t] \hat{\mathbf{h}}[t-1]\|_2^2}{\|\mathbf{x}[t]\|_2^2}, \quad (7.17)$$

and the error in the estimated graph shift operator,

$$\text{NMSE}(\mathbf{S}) = \frac{\|\mathbf{S} - \hat{\mathbf{S}}[t]\|_F^2}{\|\mathbf{S}\|_F^2}. \quad (7.18)$$

Additionally, convergence is evaluated through precision, recall, and F1 scores for the classification of non-zero elements in $\mathbf{S}$, and the missing rate (P_M) and false alarm rate (P_{FA}) of edges at steady-state are reported.

Data Generation

The algorithm was evaluated across four graph topologies: Random (R), Erdős–Rényi (ER) [Erdos, Rényi, et al., 1960], K-Regular (KR), and Stochastic Block Model (SBM) [Barabási and Albert, 1999], each with $N = 50$ nodes. The random graph weights were drawn from $\mathcal{N}(0, 1)$, thresholded between $0.3w_{\max}$ and $0.7w_{\max}$ ($w_{\max}$ being the maximum absolute weight), and normalised by 1.5 times the largest eigenvalue. For the ER graph, edges from $\mathcal{N}(0, 1)$ were thresholded between 1.6 and 1.8 in absolute value ($p_{ER} \approx 0.04$), soft-thresholded by 1.5 to range be-

tween 0.1 and 0.3, and normalised by 1.5 times the largest eigenvalue. The KR graph assigned 3 nearest-neighbour connections per node with weights from $\mathcal{U}(0.5, 1.0)$, normalised by 1.1 times the largest eigenvalue. The SBM graph comprised $k = 10$ equal-sized clusters, with intra-cluster connection probability 0.05 and inter-cluster probabilities uniform in $[0, 0.04]$. Edge weights were sampled from a Laplacian distribution with rate $\lambda_e = 2$, normalised by 1.1 times the largest eigenvalue. Following [Mei and Moura, 2016], different normalisation factors (1.1 for KR/SBM, 1.5 for Random/ER) ensure stability while preserving graph-specific properties: structured graphs require less conservative scaling than those with variable spectral properties. These factors work with filter coefficient normalisation to maintain stability in signal generation from (7.1).

Following Mei and Moura (2016), the model order is set to $P = 3$ and sparse filter coefficients h_{ij} ($2 \leq i \leq P, 0 \leq j \leq i$) are generated from $2^{i+j}h_{ij} \sim 0.5\mathcal{U}(-1, -0.45) + 0.5\mathcal{U}(0.45, 1)$, normalised by 1.5 for stability. Graph signals, $\mathbf{x}[t] \in \mathbb{R}^N$, were generated recursively from (7.1) with $\mathbf{x}[t \leq P] = \mathbf{0}$ and $\mathbf{e}[t] \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_N)$. From 11,000 generated samples, the first 1,000 were discarded to account for transitory effects and 10,000 retained for analysis. After hyperparameter optimisation, this process was repeated 5 times to construct Monte Carlo estimates of the metrics and parameters.

Estimators: ADACGP Variants and Baseline Models

Four ADACGP variants defined by two design choices were evaluated: which implementation of Algorithm 1 to use (Path 1 or Path 2), and when to apply Algorithm 2 (after Algorithm 1 reaches steady-state versus in alternation at each time step). These variants were: (1) ADACGP (P1 + debias), (2) ADACGP (P1 + alt debias), (3) ADACGP (P2 + debias), and (4) ADACGP (P2 + alt debias). For all variants, simplified instantaneous updates ($\lambda = 0$ in (7.15)) were used for $\hat{\mathbf{h}}$ to prioritise computational efficiency, as the main goal was accurate graph shift operator estimation.

Steady-state detection was used for early stopping. It utilises exponential moving averages of the observable metrics, $\text{NMSE}(\mathbf{x}_\Psi)$ and $\text{NMSE}(\mathbf{x}_h)$, denoted $\sigma[t]^{(1)}$ and $\sigma[t]^{(2)}$ respectively, where $\sigma[t] = \alpha\sigma[t-1] + (1-\alpha)\text{NMSE}[t]$ with $\alpha = 0.995$. Steady-state is reached when the moving average fails to improve by 1% within 500 epochs. Hyperparameter optimisation for each variant used a random grid search over 20,000 trials, sampling elements of $\boldsymbol{\mu}$ uniformly from $(0.001, 1]$, η from $(0.005, 0.1]$ with step 0.005, γ from $(0.05, 2.0]$ with step 0.05, and λ from $(0.80, 0.99]$ with step 0.01. The best parameters were selected by minimising the average of $\text{NMSE}(\mathbf{x}_h)$ during the final patience epochs of steady-state.

Several established baselines, both adaptive and batch methods, were benchmarked. For adaptive methods, TISO and TIRSO [Zaman, Ramos, et al., 2020] were included, sparse adaptive

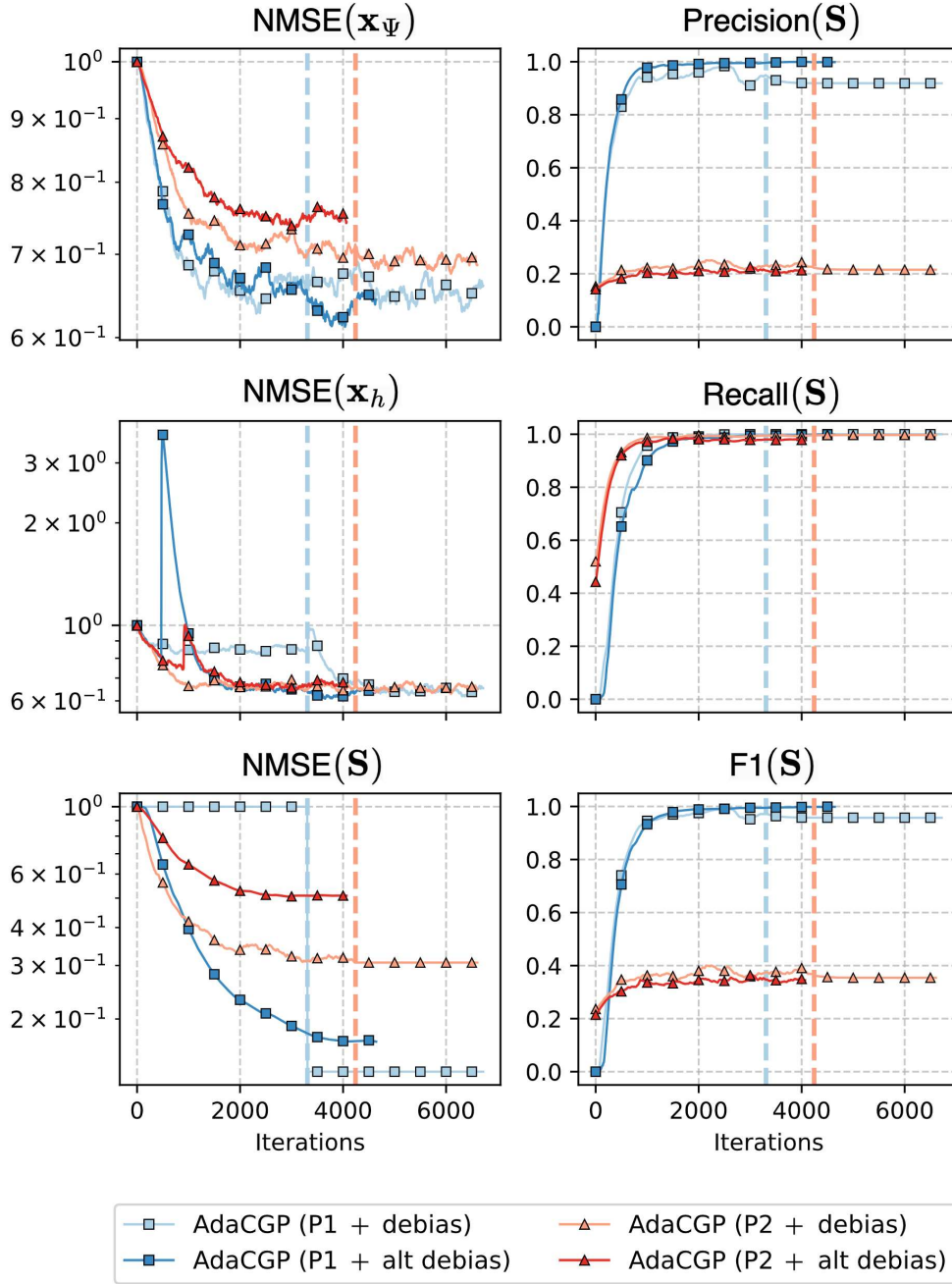


Figure 7.1: Convergence of ADACGP variants on the K-Regular topology in NMSE ($\mathbf{x}_\Psi$, $\mathbf{x}_h$, $\mathbf{S}$) and the metrics for classification of non-zero elements in $\mathbf{S}$ (Precision, Recall, F1). Vertical dashed lines indicate debiasing onset.

VAR models that track time-varying causality graphs online using LMS and RLS approaches, respectively. Unlike ADACGP, these baselines do not model structural dependencies between parameters, like shift-invariance. In these cases, graph topology estimates $\hat{\mathbf{S}}[t]$ are derived from the notion of VAR causality [Zaman, Ramos, et al., 2020], where element (i, j) is causal if $\sum_{p=1}^P \mathbb{1}(\Psi_{p,ij} \neq 0) > 0$, with weights assigned to causal non-zero edges as $\hat{\mathbf{S}}_{ij} = \|\hat{\Psi}_{ij}\|_2$. For TISO, the sparsity hyper-parameters span $(0.025, 1.0]$ with step 0.025, while TIRSO additionally samples forgetting factors from $(0.80-0.99]$ with step 0.01, optimising for minimal steady-state prediction error $\text{NMSE}(\mathbf{x}_\Psi)$. SD-SEM [Baingana and Giannakis, 2016] was also included, an adaptive method that tracks evolving topologies using structural equation models, which model observations as linear combinations of endogenous variables with sparsity assumptions. The hyper-parameter optimisation considered forgetting factors $(0.80-0.99]$ and sparsity penalties with values 1, 5, 10, 15, 20, and 25.

To ensure optimal convergence, all adaptive autoregressive models employed an adaptive step size $\alpha[t]$ (with $\mathbf{A}[t] = \alpha[t]\mathbf{I}_P$), computed as

$$\alpha[t] = \frac{2}{\lambda_{\max}(\mathbf{R}[t])} \cdot \frac{1}{\|\mathbf{x}_P[t]\|_2^2 + \varepsilon}, \quad (7.19)$$

where ε ensures numerical stability. For SD-SEM, which is not autoregressive, the default step size of $1/\lambda_{\max}$ was used. Other step-sizes $\beta[t]$ and $\rho[t]$ were calculated using the Armijo rule for automatic selection.

For batch methods, the graphical LASSO (GLasso) [Friedman, Hastie, and Tibshirani, 2007] was included, which estimates a sparse precision matrix from multivariate normal observations using the Markov property, with sparsity parameter ranging from $(0.10-1.0]$ with step 0.10. GL-SigRep [Xiaowen Dong, Thanou, et al., 2016], which learns a graph Laplacian matrix under signal smoothness assumptions, was also evaluated. GL-SigRep was applied per time-instant, and the hyper-parameter optimisation considered smoothness regularisation values 10^{-3} , $10^{-2.5}$, 10^{-2} , $10^{-1.5}$, 10^{-1} and sparsity strengths 10^{-2} , $10^{-1.5}$, 10^{-1} , $10^{-0.5}$, 10^0 . Finally, VAR and VAR+Granger [Granger, 1969] models with order $P = 3$ were included, where Granger causality determines directed edges through statistical testing of VAR model coefficients. All batch methods, except GL-SigRep, were trained with all data available up to time t . During training, the respective objective function values were monitored for convergence, and metrics were computed on the final patience epochs of steady-state. Except for VAR+Granger, where causality tests were performed on the last 50 iterations for computational efficiency.

Table 7.1: Comparison of graph topology estimation algorithms for $N = 50$. Lower values are better for all metrics. Best results are in **bold**, second best are underlined.

	Algorithm	RANDOM				ERDOS-RENYI			
		NMSE _{FC}	NMSE(S)	P _M	P _{FA}	NMSE _{FC}	NMSE(S)	P _M	P _{FA}
BATCH	GLasso	—	1.72±0.02	0.53±0.01	0.45±0.00	—	1.97±0.02	0.78±0.01	0.19±0.01
	GL-SigRep	—	1.03±0.00	0.01±0.00	0.98±0.00	—	1.04±0.00	0.01±0.00	0.98±0.00
	VAR	0.58±0.01	2.00±0.06	0.00±0.00	1.00±0.00	0.55±0.01	1.87±0.02	0.00±0.00	1.00±0.00
	VAR + Granger	0.56±0.01	1.86±0.02	0.00±0.00	0.05±0.01	0.56±0.03	1.75±0.00	0.00±0.00	0.05±0.00
ADAPTIVE	SD-SEM	—	1.40±0.01	0.55±0.00	0.44±0.00	—	1.40±0.00	0.54±0.01	0.42±0.00
	TIRSO	0.68±0.00	1.08±0.00	0.00±0.00	1.00±0.00	0.67±0.02	1.07±0.02	0.00±0.00	1.00±0.00
	TISO	0.67±0.01	1.07±0.01	0.00±0.00	1.00±0.00	0.66±0.03	1.07±0.02	0.00±0.00	1.00±0.00
	P2 + alt debias	0.53±0.00	0.07±0.02	0.00±0.00	1.00±0.00	0.59±0.01	0.23±0.01	0.02±0.00	0.05±0.01
	P2 + debias	<u>0.54±0.01</u>	0.04±0.00	0.00±0.00	1.00±0.00	0.58±0.01	0.20±0.02	0.01±0.02	0.01±0.01
	P1 + alt debias	0.56±0.01	0.07±0.00	<u>0.01±0.01</u>	<u>0.00±0.00</u>	<u>0.53±0.01</u>	<u>0.03±0.00</u>	<u>0.00±0.00</u>	<u>0.00±0.00</u>
	P1 + debias	0.56±0.00	<u>0.06±0.00</u>	0.00±0.00	0.00±0.00	0.52±0.00	0.02±0.00	0.00±0.00	0.00±0.00
	Algorithm	K-REGULAR				STOCHASTIC BLOCK MODEL			
		NMSE _{FC}	NMSE(S)	P _M	P _{FA}	NMSE _{FC}	NMSE(S)	P _M	P _{FA}
BATCH	GLasso	—	3.64±0.04	0.47±0.03	0.19±0.01	—	14.04±0.30	0.92±0.02	0.03±0.00
	GL-SigRep	—	0.91±0.00	0.14±0.00	1.00±0.00	—	1.37±0.00	0.00±0.00	0.98±0.00
	VAR	0.66±0.04	1.14±0.10	0.00±0.00	1.00±0.00	0.91±0.01	4.57±0.34	0.00±0.00	1.00±0.00
	VAR + Granger	<u>0.66±0.02</u>	0.94±0.01	0.00±0.00	0.05±0.01	0.87±0.04	3.50±0.03	0.08±0.04	0.05±0.00
ADAPTIVE	SD-SEM	—	1.41±0.00	0.61±0.01	0.43±0.01	—	3.12±0.11	0.65±0.02	0.31±0.01
	TIRSO	0.72±0.03	0.78±0.01	0.00±0.00	1.00±0.00	0.89±0.02	1.42±0.04	0.00±0.00	1.00±0.00
	TISO	0.72±0.03	0.78±0.02	0.00±0.00	1.00±0.00	0.86±0.02	1.47±0.11	0.00±0.00	1.00±0.00
	P2 + alt debias	0.76±0.01	0.52±0.02	0.02±0.00	0.60±0.03	0.78±0.00	0.26±0.03	0.03±0.01	0.91±0.01
	P2 + debias	0.70±0.01	0.33±0.02	0.01±0.00	0.44±0.15	<u>0.79±0.01</u>	0.19±0.03	0.00±0.04	0.96±0.06
	P1 + alt debias	0.67±0.02	<u>0.25±0.02</u>	<u>0.02±0.01</u>	<u>0.00±0.00</u>	0.83±0.02	<u>0.15±0.01</u>	0.09±0.02	0.24±0.06
	P1 + debias	0.66±0.01	0.13±0.02	0.00±0.00	0.01±0.02	0.84±0.01	0.09±0.17	<u>0.12±0.08</u>	<u>0.08±0.29</u>

Results

Convergence Across Graph Topologies. Fig. 7.1 shows the convergence of the metrics for the KR topology as a representative example. All models converge to low non-zero values for $\text{NMSE}(\mathbf{x}_\Psi)$ and $\text{NMSE}(\mathbf{x}_h)$, with Path 1 achieving lower errors than Path 2. The $\text{NMSE}(\mathbf{x}_h)$ for Path 1 showed significant improvement post-debiasing, and achieved near-zero $\text{NMSE}(\mathbf{S})$. While alternating debiasing converged to slightly higher values than steady-state debiasing, it demonstrated faster convergence in $\text{NMSE}(\mathbf{S})$ for Path 1.

The considered methods exhibited distinct differences in $\text{Precision}(\mathbf{S})$, with Path 1 maintaining near-perfect precision (≈ 1.0) compared to Path 2’s plateau around 0.2. All variants achieved high $\text{Recall}(\mathbf{S})$ (≈ 1.0), though Path 2 showed faster initial increases. The F1 scores strongly favour Path 1 (≈ 1.0 versus ≈ 0.4), with the alternating debiasing variant showing marginally better precision due to its iteration-by-iteration approach providing greater robustness to noise and misspecification of non-zero elements upon entering Algorithm 2.

As shown in Table 7.1, the ADACGP variants consistently outperformed the baselines across all

graph topologies. To facilitate comparison across algorithms, NMSE_{FC} is included in Table 7.1, representing the forecast error of each model (if available), which equals $\text{NMSE}(\mathbf{x}_h)$ for ADACGP variants.

Specifically, ADACGP (P1 + debias) achieves the lowest or second-lowest $\text{NMSE}(\mathbf{S})$ across all topologies, with improvements in excess of 94%, 98%, 83%, and 93% for Random, Erdos-Renyi, K-Regular, and SBM topologies, respectively, compared to baseline models. Among adaptive algorithms, TISO and TIRSO fail to enforce sparsity ($P_{FA} = 1.00$ across all topologies), while SD-SEM sets some edges to zero but with consistently high error rates. In contrast, ADACGP Path 1 variants maintain near-zero false alarm rates ($P_{FA} \leq 0.08$) while achieving minimal missing edge rates ($P_M \leq 0.12$) across all topologies. These results highlight that baseline methods struggle with their sparsity mechanisms' inability to explicitly set elements to zero, while the variable splitting approach overcomes this limitation, effectively identifying and removing irrelevant edges for an improved P_{FA} and P_M trade-off. While for batch methods, GLasso achieves lower P_{FA} values but at the expense of significantly higher P_M .

The SBM topology presents a particularly challenging case due to its high sparsity ($\approx 1\%$ non-zero elements). Here, ADACGP (P1 + debias) is surpassed only by VAR+Granger in P_M (0.12 ± 0.08 vs. 0.08 ± 0.04) and P_{FA} (0.08 ± 0.29 vs. 0.05 ± 0.00). The higher variance in P_{FA} results from debiasing being applied upon convergence when facing extremely sparse topologies, where minor edge detection errors significantly impact metrics. However, despite methodological differences between VAR+Granger (sparsity obtained by offline statistical testing) and ADACGP (sparsity obtained online), ADACGP (P1 + debias) achieves a $\approx 39\times$ reduction in $\text{NMSE}(\mathbf{S})$ (0.09 ± 0.17 vs. 3.50 ± 0.03).

It is important to note that $\text{NMSE}(\mathbf{x}_h)$ and $\text{NMSE}(\mathbf{S})$ evaluate different aspects of model performance (prediction accuracy vs. topology estimation accuracy). Due to the structure of the alternating optimisation framework in (7.4)-(7.6), minimising $\text{NMSE}(\mathbf{x}_h)$ often leads to improved $\text{NMSE}(\mathbf{S})$, while the converse is not necessarily true.

Identification of the Graph Shift Operator Topology. Fig. 7.2 compares the true and estimated graph shift operator matrices across the R, ER, KR, and SBM topologies. For each topology, the true topology, $\mathbf{S}$, is shown alongside the mean estimate over 20 realisations, $\text{Mean}(\hat{\mathbf{S}})$, from the best-performing ADACGP variant (based on lowest steady-state $\text{NMSE}(\mathbf{x}_h)$). While there is some bias in the estimated element values, particularly in the KR topology, the relative edge weights and overall structure were well preserved.

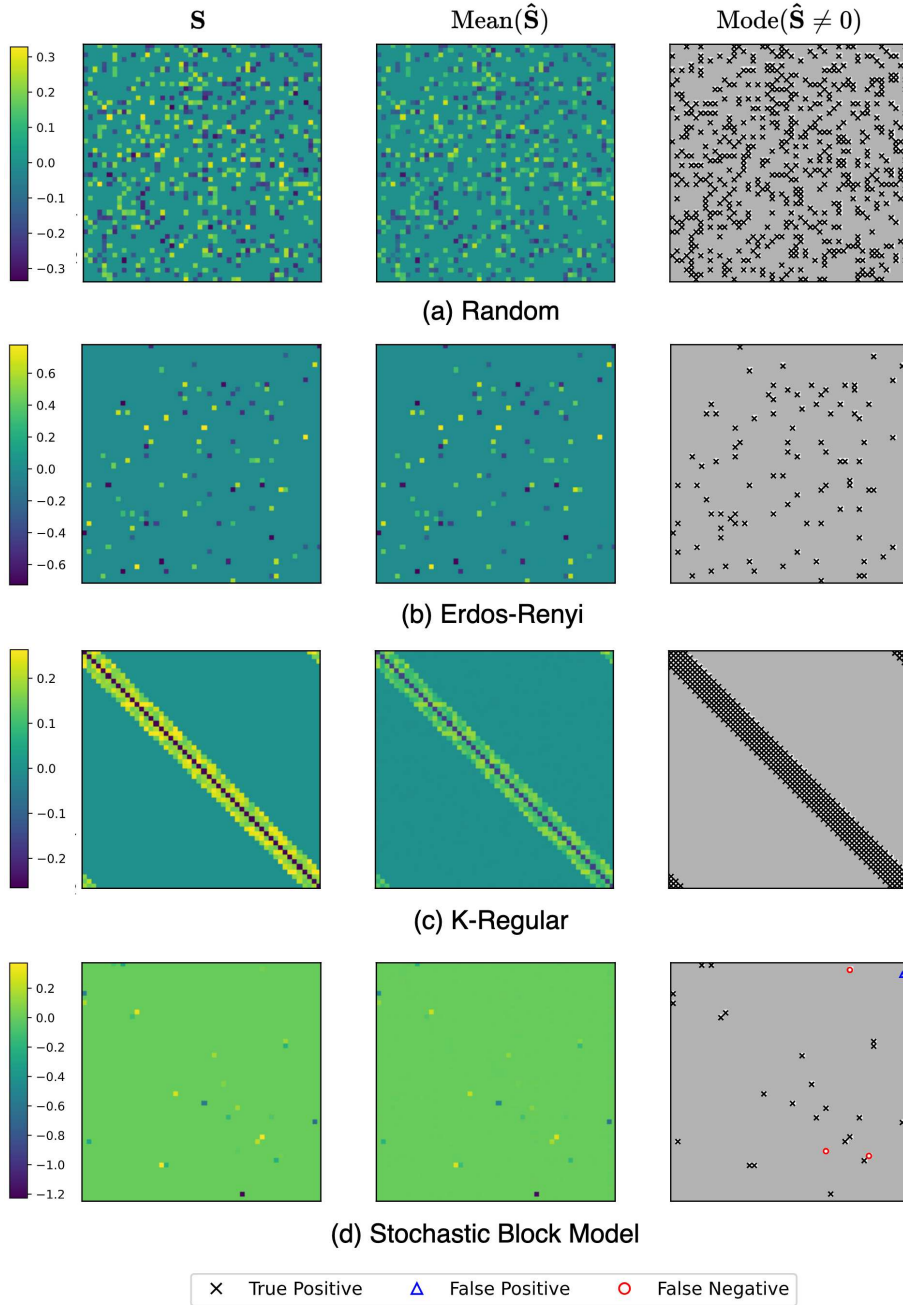


Figure 7.2: Graph shift operator matrices across four graph topologies: (a) Random, (b) Erdős–Rényi, (c) K-Regular, and (d) Stochastic Block Model. For each topology: true $\mathbf{S}$ (left); mean estimate over 20 realisations, $\text{Mean}(\hat{\mathbf{S}})$ (middle); and most commonly estimated non-zero elements, $\text{Mode}(\hat{\mathbf{S}} \neq 0)$, with highlighted true positives and false negatives (right). Results shown for the best-performing ADACGP variant based on steady-state $\text{NMSE}(\mathbf{x}_h)$.

A comparison of the most common non-zero elements over realisations, $\text{Mode}(\hat{\mathbf{S}} \neq 0)$, to the true elements in $\mathbf{S}$ reveals minimal false positives and false negatives. This demonstrates the effectiveness of ADACGP’s sparse algorithm in explicitly zeroing out GSO matrix elements while identifying most of the true edges.

The Effect of Sparsity Regularisation on Performance. To investigate how estimated network sparsity relates to $\text{NMSE}(\mathbf{x}_h)$, a practical metric for hyper-parameter tuning that does not require knowledge of the true network structure, experiments were conducted with varying sparsity regularisation strength. Using the best-performing hyper-parameters from the ADACGP variants from before, elements of $\boldsymbol{\mu}$ were randomly sampled uniformly from $(0.001, 1]$ across 5,000 trials per model variant and different graph types with $N = 50$ nodes.

Fig. 7.3 shows the $\text{NMSE}(\mathbf{x}_h)$ versus non-zero elements in $\hat{\mathbf{S}}$ (left column) and P_M versus P_{FA} trade-offs (right column), where the x-axis has been uniformly binned and median results displayed. As expected, increasing sparsity regularisation reduces the estimated number of non-zero elements and increases P_M . Path 1 variants minimised $\text{NMSE}(\mathbf{x}_h)$ near the true sparsity level, while Path 2 variants overestimated the number of non-zero elements. Path 1 consistently achieved lower $\text{NMSE}(\mathbf{x}_h)$ and superior P_M – P_{FA} trade-offs across all topologies, exhibiting near-perfect performance with very low P_M and P_{FA} . These results demonstrate that optimising $\text{NMSE}(\mathbf{x}_h)$ for Path 1 variants leads to accurate network sparsity estimation and thus improved causal discovery. It is noted that the results in this section serve a different purpose than those in Tab. 7.3.1 and cannot be directly compared. While the previous analysis used optimised hyper-parameters, this section deliberately varies regularisation strength to explore the relationship between sparsity and performance metrics, with results aggregated across multiple configurations including sub-optimal ones for improved visualisation.

Comparison of Runtime and Resource Utilisation. To assess computational efficiency, CPU time per iteration and peak memory usage for the Random topology across network sizes ($N = 10$ to 2500) were measured, averaging results over three independent realisations. All algorithms were implemented in NUMPY [Harris et al., 2020] on an AMD EPYC 7513 CPU, and original source code for baselines was used when available. For all auto-regressive methods, $P = 3$ was set, and all adaptive algorithms employed a fixed step-size. Fig. 7.4 illustrates these metrics for all algorithms, with empirical scaling exponents derived by fitting polynomial regression models to the highest 50% of tested N values. ADACGP’s debiasing step (Algorithm 2) was excluded from these experiments, since it is optional and has the same worst-case complexity as Algorithm 1.

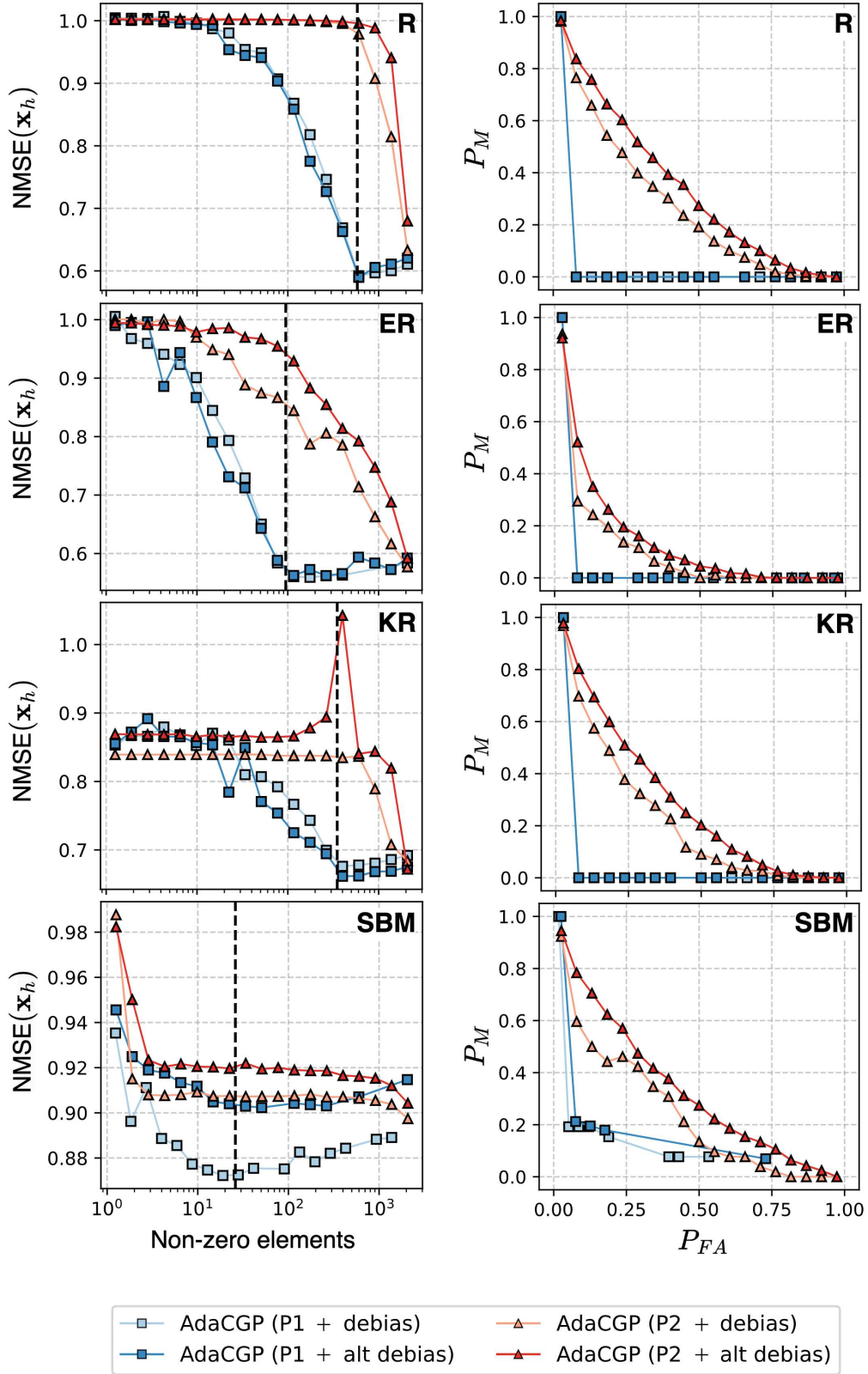


Figure 7.3: Performance of ADACGP variants for different graph topologies: R, ER, KR, and SBM. Left column: $\text{NMSE}(\mathbf{x}_h)$ vs. estimated non-zeros in $\hat{\mathbf{S}}$. Vertical dashed lines indicate the true number of non-zeros. Right column: probability of missing edges (P_M) vs. probability of false alarm (P_{FA}).

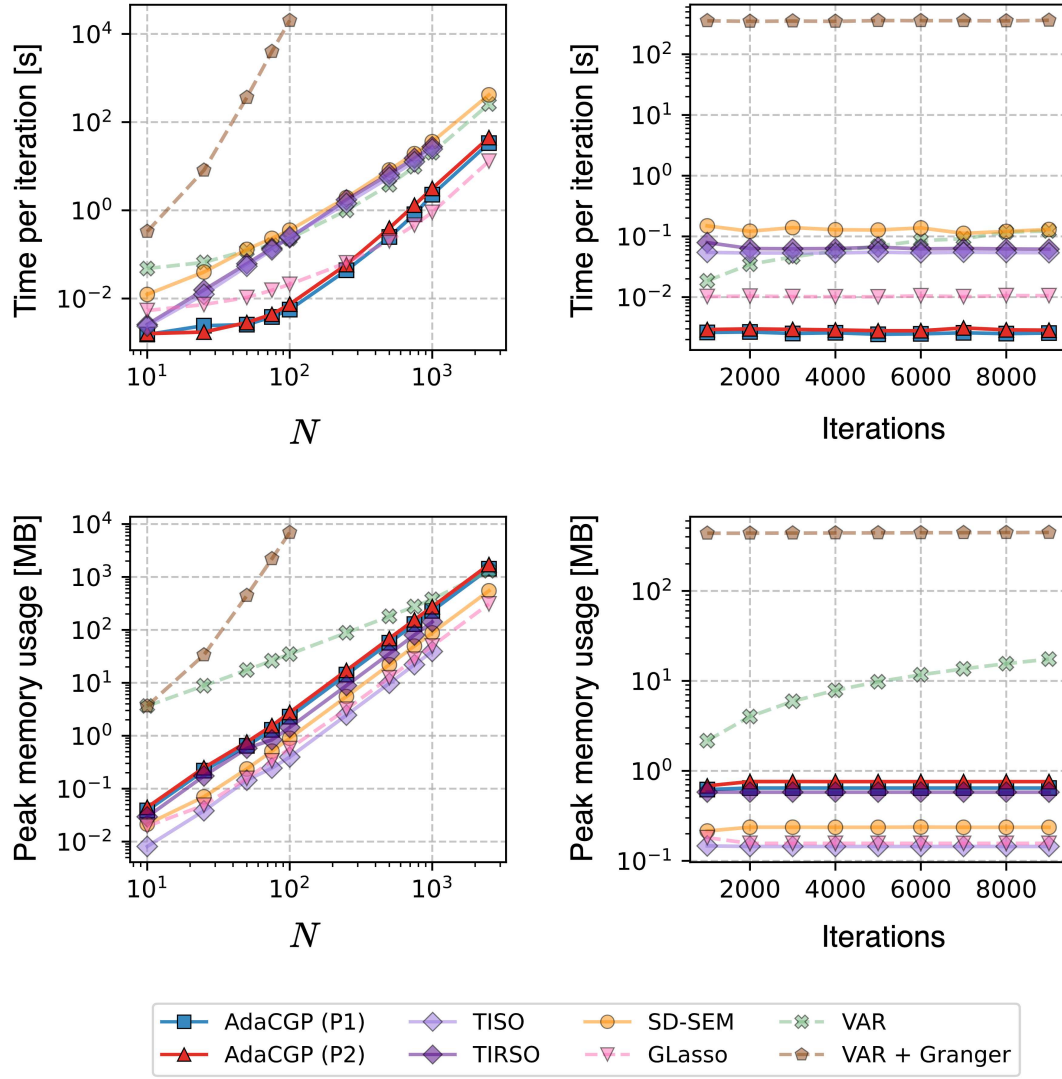


Figure 7.4: Computational complexity analysis of graph topology learning algorithms. Left panels: scaling of time (top) and memory usage (bottom) with network size, N . Right panels: time (top) and memory usage (bottom) as a function of iterations.

Table 7.2: Scaling of per iteration time and memory complexity. Asterisk denotes theoretical scaling.

	Algorithm	COMPLEXITY	
		Time	Memory
BATCH	GLasso	$O(N^{2.31})$	$O(N^{1.99})$
	GL-SigRep	$O(N^{2.00})^*$	—
	VAR	$O(N^{2.42})$	$O(N^{1.17})$
	VAR + Granger	$O(N^{5.81})$	$O(N^{3.96})$
ADAPTIVE	SD-SEM	$O(N^{2.32})$	$O(N^{2.00})$
	TIRSO	$O(N^{2.05})$	$O(N^{2.00})$
	TISO	$O(N^{1.98})$	$O(N^{2.00})$
	AdaCGP (P2)	$O(N^{2.89})$	$O(N^{2.00})$
	AdaCGP (P1)	$O(N^{2.91})$	$O(N^{2.00})$

The left panels of Fig. 7.4 show how the computational complexity of each algorithm scales with N . ADACGP variants demonstrate empirical time complexity of $O(N^{2.91})$ for Path 1 and $O(N^{2.89})$ for Path 2, aligning with the theoretical $O(N^3)$ complexity in Sec. 7.2.3. Despite cubic scaling, the implementation uses efficient matrix operations, achieving competitive runtimes across all tested networks ($N = 10$ –2500) and outperforming several baselines. Memory usage follows similar patterns, with ADACGP variants maintaining $O(N^{2.00})$ scaling comparable to adaptive baselines. Table 7.2 summarises the scaling properties for all algorithms, with VAR+Granger showing highest time complexity, $O(N^{5.81})$, due to causality testing across all potential edges. For GL-SigRep, the theoretical time complexity of $O(N^{2.00})$ from [Xiaowen Dong, Thanou, et al., 2016] is reported, as the implementation was not sufficiently optimised for large-scale benchmarking. Additionally, TIRSO’s time complexity of $O(N^{2.05})$ is observed to be significantly lower than its theoretical scaling of $O(N^{3.00})$, likely because the cubic scaling has not become dominant at the network sizes tested.

The right panels of Fig. 7.4 demonstrate the constant per-iteration time and memory complexity of adaptive methods, a crucial property that renders them advantageous for the efficient analysis of streaming signals. In contrast, batch methods such as VAR show increasing computational demands as the training window grows (this observation is less apparent for VAR+Granger due to the significant overall cost). As discussed in Sec. 7.2.3, sparsity could be leveraged to yield further improvements in scalability through an implementation of ADACGP using optimised sparse matrix libraries.

7.3.2 Assessing the Structure and Stability of Cardiac Fibrillation

Having validated ADACGP’s performance on synthetic data, the algorithm is now applied to its motivating clinical application: characterising the functional connectivity of cardiac fibrillation. While existing applications of Granger causality to fibrillation dynamics [Handa, X. Li, et al., 2020] use batch processing to capture average propagation patterns, ADACGP tracks temporal changes in graph shift operator weights and topology. This enables assessment of propagation patterns and their stability at both global and local levels, with the potential to inform clinical evaluations of the electrophenotype (Sec. 3.3), where propagation appears chaotic yet maintains quantifiable structure [Petrutiu, J. Ng, et al., 2006; Sanchez-Munoz, Rojo-Alvarez, et al., 2009; Handa, X. Li, et al., 2021].

Dataset

Optical mapping data of VF from an *ex vivo* Sprague-Dawley rat heart during sequential administration of rotigaptide at concentrations in micromolar (μM), an anti-arrhythmic drug shown to induce more organised dynamics [Handa, X. Li, et al., 2021], were analysed. All procedures were performed in accordance with the UK Animals (Scientific Procedures) Act 1986 and ARRIVE guidelines, and were approved by the Imperial College London Ethical Review Board (project licences PEE7C76CD and PCA5EE967). Animal procedures conformed to guidelines from Directive 2010/63/EU of the European Parliament on the protection of animals used for scientific purposes. Rats were anaesthetised with 5% isoflurane (95% oxygen mix) and euthanised via cervical dislocation prior to heart explantation.

The experiment records fluorescence signals proportional to the cardiac action potential (Sec. 3.1) using a 128×80 pixel camera at 1000 Hz. Recordings span 10 s during administration of rotigaptide at concentrations of 0, 10, 30, 50, and 80 μM , where zero represents baseline VF. The pre-processing pipeline followed established standards for cardiac optical mapping data [Laughner, F. S. Ng, et al., 2012; Handa, X. Li, et al., 2021]: spatial filtering (3×3 uniform kernel), temporal filtering (0–100 Hz low-pass), baseline drift removal, and amplitude normalisation. Phase angles were calculated from the Hilbert transform. The processed data was spatially downsampled by a factor of 4 to ≈ 350 nodes, with mean-standard-deviation normalisation applied to form the graph signals, $\mathbf{x}[t]$. For the graph-based methods of this thesis, the pixel locations are treated as graph nodes on a two-dimensional regular grid, with edges connecting spatially adjacent pixels (Sec. 3.5.2).

Estimators

ADACGP (P1 + alt debias) is deployed with gradients constrained to adjacent pixels to reflect the local connectivity of cardiac tissue. The model uses parameters $P = 3$, $\gamma = 1$, $\eta = 0.01$, elements of $\boldsymbol{\mu}$ as 0.1, and $\lambda = 0.5$, with low forgetting factor enabling rapid adaptation. ADACGP (P2 + alt debias) uses identical hyper-parameters, differing only in the algorithmic approach as detailed in Section 7.2.3. Step-sizes for both variants were calculated using the Armijo rule for automatic selection. To benchmark the approach against established methods in the cardiac electrophysiology literature, VAR+Granger was implemented as a comparative baseline, which has been widely used for functional connectivity estimation in cardiac fibrillation [Handa, X. Li, et al., 2020; Alcaine, Mase, et al., 2016]. Unlike the online ADACGP approach, this offline method was trained on the entire time series data. Individual VAR models ($P = 3$) were trained for each pixel, using only time series from adjacent pixels to reflect local connectivity and reduce computational complexity. Granger causality tests were then performed on the resulting VAR coefficients at each pixel to determine significant causal connections.

Results

Fig. 7.5 demonstrates local stability through edge patterns at the start (t^*) and end (T^*) of steady-state. During VF ($0 \mu\text{M}$), the magnitude and direction of net edges varied considerably between t^* and T^* , indicating unstable propagation patterns. The highlighted region's Out-In Degree ($\sum_j \hat{S}_{ij} - \sum_i \hat{S}_{ij}$) fluctuated significantly between positive (source) and negative (sink) values, demonstrating this instability. This dynamic behaviour is captured differently by the algorithm variants, with ADACGP (P1 + alt debias) showing more rapid tracking of changes compared to ADACGP (P2 + alt debias), due to the removal of commutativity regularisation in Path 1. In contrast, at $50 \mu\text{M}$, edge magnitudes and directions across time points remained largely consistent, suggesting more stable propagation patterns, with the highlighted region maintaining consistent source characteristics and minimal Out-In Degree variations.

When compared with VAR + Granger (Fig. 7.5), the advantages of the online approach are evident. The offline nature of Granger causality, which analyses the entire time series at once, only captures average propagation patterns and misses the temporal dynamics. This limitation is particularly pronounced during baseline VF ($0 \mu\text{M}$), where Granger identifies very few causal edges due to the highly variable dynamics that get averaged out. In contrast, ADACGP captures these dynamical variations as evidenced by the changing graph topology between t^* and T^* . While Granger could theoretically be applied in a sliding window approach, the computational complexity of retraining

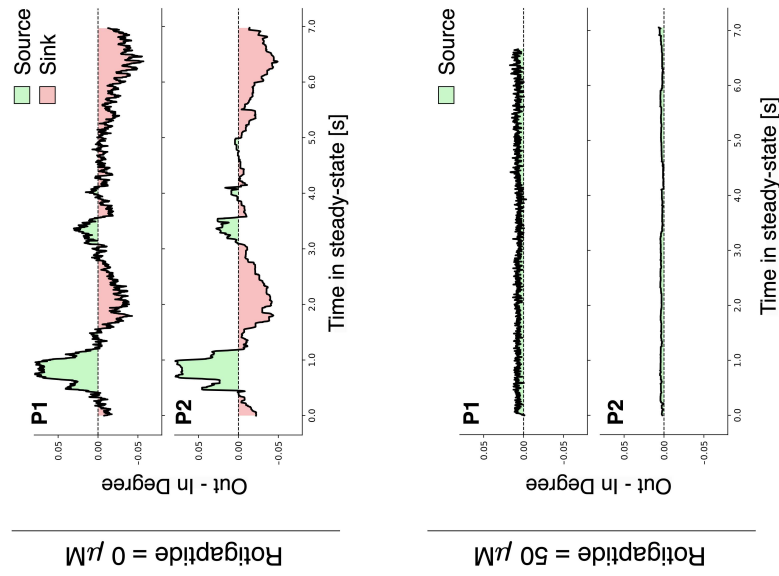


Figure 7.5: Local stability analysis of fibrillation dynamics. Net edge directions at steady-state start ($t = t^*$) and end ($t = T^*$) for VF ($0 \mu\text{M}$, top row) and rotating $50 \mu\text{M}$ (bottom row) using ADACGP (P1). Left panels show Out-In Degree time series for circled regions comparing P1 and P2 variants, demonstrating alternating source-sink behaviour in VF versus stable source characteristics at $50 \mu\text{M}$. Right panels present VAR + Granger comparison. Colour and arrow length denote edge weights.

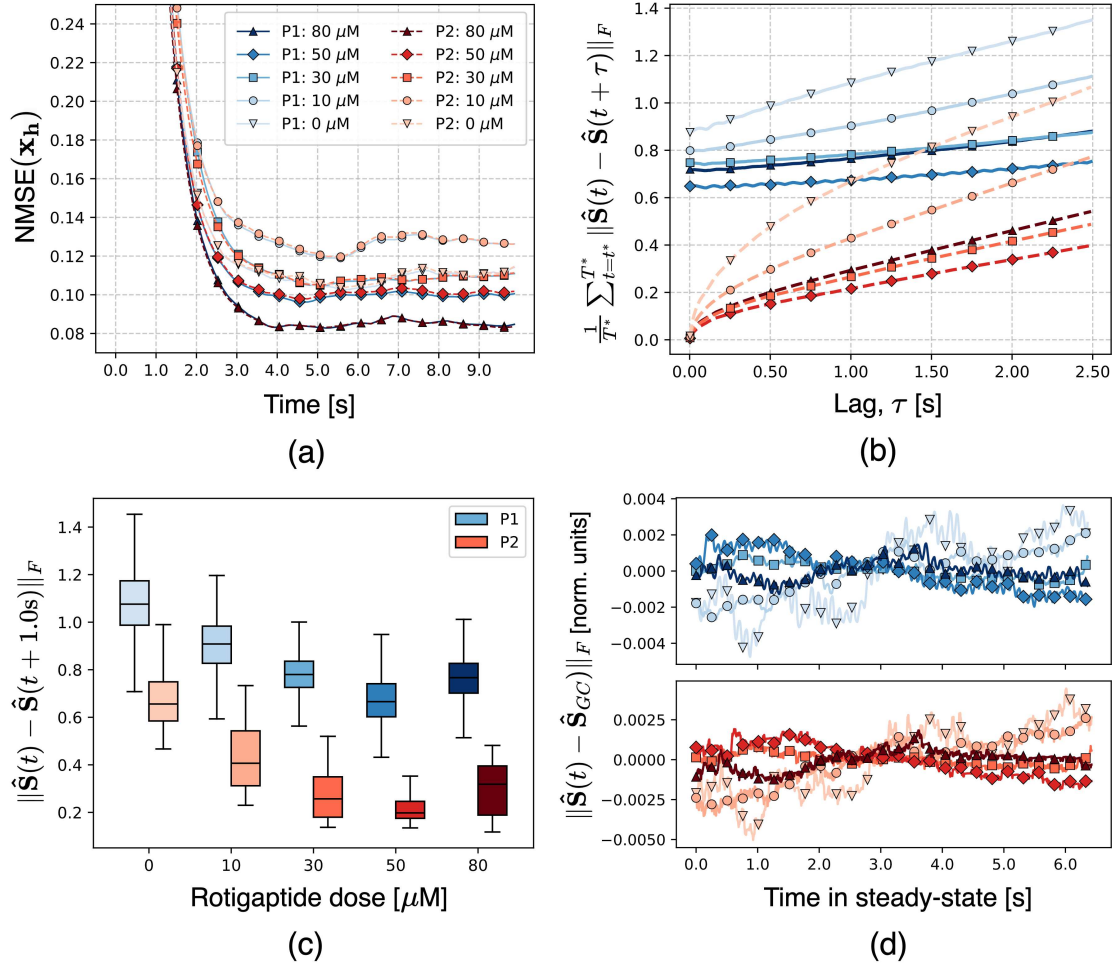


Figure 7.6: Global analysis of fibrillation dynamics across varying rotigaptide concentrations. (a) Prediction error $NMSE(\mathbf{x}_h)$ over time, showing lower steady-state errors with increasing concentrations. (b) Time-lag analysis of Frobenius norm differences between graph shift operators, demonstrating more stable topology estimates at higher concentrations and faster adaptation of ADACGP P1. (c) Distribution of Frobenius norms between graph shift operators lagged by 1.0 second during steady-state, indicating more stable dynamics at higher concentrations. (d) Temporal evolution of differences between ADACGP and VAR + Granger topologies, showing decreased variation with increasing rotigaptide concentration.

for graphs of this size ($\sim 10^2$ nodes) makes real-time analysis of cardiac data streams impractical, further highlighting the value of the adaptive approach.

Fig. 7.6 reveals key aspects of global stability. First, $\text{NMSE}(\mathbf{x}_h)$ (Fig. 7.6a) showed lower steady-state errors with increasing rotigaptide concentrations for both ADACGP variants, suggesting more organised dynamics consistent with the drug’s anti-arrhythmic nature [Hsieh, Lin, et al., 2016]. Second, Fig. 7.6c showed Frobenius norms between successive graph shift operators exhibiting lower median values and variability at higher concentrations, indicating more stable network topologies. The time-lag analysis (Fig. 7.6b) revealed important methodological differences: while both variants showed similar error patterns at larger time scales (> 1 second), ADACGP (P1) maintained finite error at short time lags (< 1 second) while ADACGP (P2) approached zero, confirming Path 1’s greater sensitivity to topology changes.

Comparison with VAR + Granger (Fig. 7.6d) showed high temporal variation between ADACGP’s dynamic graphs and Granger’s static graph during baseline VF ($0 \mu\text{M}$), with these variations diminishing at higher rotigaptide concentrations. This demonstrates how ADACGP captures significant dynamic changes in graph structure that are averaged out in Granger’s time-invariant representation, while confirming that both approaches align when the underlying dynamics become more stable.

7.4 Discussion

This chapter introduced ADACGP, an online time-vertex adaptive filtering algorithm for tracking time-varying causal graph topologies from streaming multivariate time series, addressing **RQ3** by demonstrating that sparse, evolving graph topologies can be reliably learned and tracked in non-stationary systems. The variable splitting approach for sparsity-aware adaptive filtering efficiently identifies the graph shift operator in real time, reliably estimating its causal elements. Experiments across diverse graph topologies demonstrated ADACGP’s superior performance over existing adaptive VAR models in both prediction accuracy and topology recovery. Path 1 variants showed particularly strong performance in causal discovery while maintaining near-zero false positive rates, with prediction error minimisation proving effective for hyperparameter tuning and sparsity estimation. Application to rat VF data revealed ADACGP’s ability to track changes in cardiac organisation under varying rotigaptide concentrations, providing insights into the structure and stability of fibrillation at both global and local levels.

Variable splitting and sparsity. The central methodological contribution of this chapter is the variable splitting approach for achieving true sparsity in online estimation setting. Existing adaptive methods for sparse topology learning, including TISO, TIRSO [Zaman, Ramos, et al., 2020], and SD-SEM [Baingana and Giannakis, 2016], employ ℓ_1 -regularisation but consistently fail to drive non-causal elements to exact zero, as evidenced by their $P_{FA} = 1.00$ across all topologies tested (Tab. 7.1). For topology estimation, the binary presence or absence of an edge carries causal significance that cannot be recovered by thresholding small-but-nonzero values. By decomposing each variable into non-negative components and solving through projected gradient descent, the variable splitting formulation converts the ℓ_1 penalty into a hard non-negativity constraint that produces exact zeros by design.

Path 1 versus Path 2. The two algorithmic paths represent different trade-offs between computational cost and estimation quality. Path 1 defers the enforcement of shift-invariance to the $\hat{\mathbf{S}}$ update step, keeping the $\hat{\Psi}$ sub-problem convex, and solves a separate regularised least squares problem to obtain a debiased estimate of $\mathbf{S}$. Path 2 enforces commutativity directly in the $\hat{\Psi}$ update and approximates $\hat{\mathbf{S}} = \hat{\Psi}_1[t]$, avoiding the additional optimisation at the cost of bias. The experiments consistently favour Path 1: it achieves near-perfect precision (≈ 1.0 versus ≈ 0.2 for Path 2), substantially lower $\text{NMSE}(\mathbf{S})$, and crucially, its prediction error $\text{NMSE}(\mathbf{x}_h)$ is minimised near the true sparsity level (Fig. 7.3), making it a reliable proxy for hyperparameter selection without knowledge of the ground truth topology. In the cardiac application, Path 1 also demonstrated greater sensitivity to short-timescale topology changes (Fig. 7.6b), confirming that the additional computational cost translates into practical benefits for tracking non-stationary systems. More broadly, the forgetting factor λ is the mechanism through which ADACGP operationalises non-stationarity. By exponentially weighting past observations, it allows the algorithm to track gradual topology changes without assuming stationarity of the underlying process, directly addressing the core requirement of **RQ3**.

Clinical implications for cardiac fibrillation. The application to rat VF data demonstrated that ADACGP captures dynamic changes in cardiac organisation that batch Granger causality misses. During baseline VF, where Granger identifies very few causal edges because the highly variable dynamics are averaged out, ADACGP tracks the evolving topology and reveals alternating source-sink behaviour at the local level (Fig. 7.5). The convergence of ADACGP’s dynamic estimates toward Granger’s static estimate at higher rotigaptide concentrations (Fig. 7.6d) provides mutual validation: both methods agree when the underlying dynamics are stable, but only the adaptive approach captures the temporal structure when they are not. This dynamic assessment of functional

connectivity could inform clinical treatment by identifying the stability of conduction patterns maintaining the arrhythmia and highlighting critical areas of the myocardium for radio-frequency catheter ablation (Sec. 3.4). Such capability is particularly valuable in redo ablation cases, where scar tissue from previous procedures distorts the cardiac anatomy and alters myocardial electrical properties [Kyriakou, Thomas, et al., 2023].

Broader applications. Beyond cardiac electrophysiology, ADACGP is applicable to any domain where functional connectivity evolves over time. In neuroscience, the approach could track dynamic connectivity in the brain, improving the characterisation of conditions such as schizophrenia, Alzheimer’s disease, and depression through the identification of abnormal transient connectivity patterns [Hasan, Pandey, and Khan, 2021; Alteriis, Sherwood, et al., 2025]. It could also provide insights into how brain networks reconfigure during cognitive tasks such as memorisation, learning, and decision-making [Preti, Bolton, and Van De Ville, 2017]. In finance, the framework could track shifting causal dependencies among assets during market regime changes [Papana et al., 2017], and in infrastructure networks such as power grids, it could monitor evolving propagation patterns in real time [Ahmad and Papadopoulos, 2022].

Limitations and future directions. For clinical translation, several challenges remain: validation in human subjects with appropriate ground truth comparisons, systematic characterisation of performance under the noise conditions typical of clinical recordings, and the development of clinically interpretable outputs. From a methodological perspective, future work could extend the framework to non-linear adaptive filtering to better capture the chaotic dynamics of fibrillation, and incorporate domain-specific graph shift operator constraints, such as the anisotropic Laplacian or energy-preserving shift operators [Scalzo et al., 2023], through the variable splitting approach, aligning more closely with established physical models of cardiac conduction [F. Fenton and Karma, 1998].

7.5 Summary and Relation to Thesis

This chapter directly addresses the third research question of this thesis (**RQ3**): *how can sparse, time-varying graph topologies be learned and tracked from streaming multivariate time series in non-stationary systems?* ADACGP instantiates the CGP model of Sec. 2.3.1 within the adaptive filtering framework of Sec. 2.1.4, using the variable splitting approach to achieve the true sparsity that existing online methods cannot. The result is a principled and efficient framework for tracking evolving causal relationships from streaming data, bridging the gap between GSP theory

and adaptive signal processing. Unlike FIBMAP, which is a task-specific architecture tailored to the challenges of AF mapping, ADACGP is a general-purpose algorithm applicable to any domain requiring real-time estimation of functional connectivity from non-stationary multivariate time series.

Within the broader thesis, ADACGP operates on fully observed spatiotemporal signals at each node, and complements FIBMAP (Chapter 5) in the progressive *reconstruct–analyse–characterise* pipeline. FIBMAP *reconstructs* global dynamics from sparse sequential catheter measurements, producing dense spatiotemporal signals across the entire atrial surface. COREL (Chapter 6) *analyses* the reliability of these reconstructions through calibrated prediction intervals. ADACGP *characterises* the underlying fibrillation mechanism by learning the time-varying directed functional connectivity from these signals, identifying conduction pathways and assessing their temporal stability as a basis for ablation planning.

Chapter 8

Conclusion and Future Directions

8.1 Conclusion

This thesis addressed the problem of learning from spatiotemporal signals on irregular graph domains, with a particular focus on advancing the methodological foundations of graph-based spatiotemporal machine learning and demonstrating their application to cardiac electrophysiology. The thesis was structured around a progressive pipeline, *reconstruct–analyse–characterise*, designed to address the clinical challenge of personalised catheter ablation of atrial fibrillation: first recovering global dynamics from sparse measurements, then quantifying the reliability of those predictions, and finally identifying the fibrillation mechanism sustaining the arrhythmia. While the individual components have been validated independently and patent applications have been filed to protect the imputation mapping methodology, full integration and prospective clinical validation remain as future work. The contributions and findings of each chapter are summarised below, before discussing their broader implications and directions for future research.

Chapter 4 introduced PULSEDIFF, a template-guided denoising diffusion probabilistic model for imputing missing values in ECG signals recorded by wearable devices. By extracting a subject-level waveform template from the observed signal components, applying beat-level stochastic augmentation to capture morphological variability, and adapting the prior strength through a rhythm confidence score, PULSEDIFF improved the reconstruction accuracy of two baseline diffusion architectures and outperformed the leading deterministic method for transient missing data on the PTB-XL benchmark. Beyond its standalone contributions, this chapter established both the strengths and the boundaries of temporal priors for physiological signal reconstruction: while quasi-periodic structure provides a powerful inductive bias for body-surface ECG signals, the chaotic spatiotemporal dynamics of cardiac fibrillation demand additional spatial structure, motivating the graph-based representations adopted in subsequent chapters.

Chapter 5 presented FIBMAP, the central contribution of this thesis, directly addressing **Research Question 1**. FIBMAP formulated cardiac mapping as a spatiotemporal graph signal reconstruction problem and demonstrated that a graph recurrent neural network, combining anisotropic message

passing with recurrent temporal processing and a hybrid global-local embedding strategy, can reconstruct global fibrillation dynamics from only 10% instantaneous surface coverage. Validated on high-resolution optical mapping of rat ventricular fibrillation (17 hearts, 253 million observations) and clinical non-contact mapping of human atrial fibrillation (51 patients, 4 billion observations), FIBMAP outperformed baselines by 47–53% in mean absolute error and 23–109% in the recovery of phase singularity locations. A supplemental translational showcase demonstrated that personalised reconstruction from routine contact mapping recordings is feasible, establishing imputation mapping as a new modality for visualising global AF dynamics from sparse routine recordings. By recovering global dynamics that are inaccessible through conventional sequential contact mapping, this approach takes a concrete step toward electrophenotype-informed AF ablation strategies beyond empirical pulmonary vein isolation. Patent applications have been filed to protect the intellectual property associated with this approach.

Chapter 6 introduced COREL, addressing **Research Question 2** by developing a distribution-free conformal prediction framework that exploits relational structure among correlated time series to provide calibrated uncertainty quantification for spatiotemporal forecasts. By training a spatiotemporal graph neural network on prediction residuals from a calibration set, COREL constructs prediction intervals that account for inter-series dependencies without requiring modification of the base forecasting model. The method learns a sparse dependency graph directly from the residuals and incorporates an adaptive component to handle non-stationary error distributions. Applied to traffic, air quality, and energy consumption benchmarks, COREL achieved state-of-the-art uncertainty quantification with accurate coverage guarantees, producing tighter prediction intervals than methods that treat each time series independently.

Chapter 7 developed ADACGP, addressing **Research Question 3** through a sparsity-aware adaptive algorithm for learning time-varying graph topologies from streaming multivariate time series. By extending the causal graph process framework to the online setting through a variable splitting approach, ADACGP achieves true sparsity in estimated topologies, overcoming a fundamental limitation of existing adaptive methods that fail to drive non-causal elements to exact zero. Through comprehensive simulations, ADACGP achieved improvements exceeding 83% in topology estimation compared to existing methods. Applied to *ex vivo* optical mapping recordings of rat ventricular fibrillation, ADACGP tracked dynamic propagation patterns more effectively than batch Granger causality, identifying temporal variations in functional connectivity that static approaches average out. The algorithm characterises the temporal stability of conduction patterns, a property directly relevant to distinguishing sustained arrhythmic drivers from transient wavefront activity.

Together, these contributions establish a comprehensive framework for graph-based spatiotemporal machine learning applied to cardiac electrophysiology. FIBMAP represents a task-specific architecture tailored to the unique challenges of AF mapping, demonstrating the clinical viability of graph-based spatiotemporal approaches in this domain. In contrast, COREL and ADACGP are general-purpose methodologies applicable across diverse spatiotemporal problems beyond cardiac electrophysiology, from traffic forecasting and air quality monitoring to neuroscience and financial network analysis. While the underlying principles and algorithms extend to any domain requiring the analysis of time-varying signals on irregular network structures, the clinical impact is already tangible. FIBMAP enables global AF dynamics to be reconstructed from sparse clinical recordings, and ADACGP tracks dynamic functional connectivity from fibrillation data where batch methods fail, both validated on real cardiac tissue and representing capabilities that did not exist prior to this thesis.

8.2 Future Directions

Several promising directions for future research are identified, organised into methodological extensions and clinical translation.

8.2.1 Methodological Extensions

Improving reconstruction fidelity and scalability. As discussed in [Sec. 5.6](#), the translational showcase demonstrated that pre-trained models generalise across recording modalities via fine-tuning, suggesting that FIBMAP is not tightly bound to any single training source. Fusing diverse data sources into a unified training framework, combining high-resolution optical mapping, large-scale clinical non-contact recordings, and biophysical simulations, could yield a stronger predictive model than any single modality alone [[Y. Huang et al., 2021](#)], and the optimal composition of such multi-modal training data remains an open question. A further consideration is scalability: FIBMAP’s message-passing operations scale superlinearly in the number of graph nodes, limiting the spatial resolution at which reconstruction is computationally feasible. Promising avenues include architectures based on ordinary differential equations that enable non-dissipative long-range propagation through stable dynamical systems [[Gravina et al., 2024](#); [Trenta, Gravina, and Bacciu, 2025](#); [Chamberlain et al., 2021](#)], graph transformer architectures [[Dwivedi and Bresson, 2021](#); [Müller et al., 2024](#)], and hierarchical graph representations that enable multi-resolution information flow [[R. Lam et al., 2023](#); [Cini, Marisca, Bianchi, et al., 2023](#)].

Integration of reconstruction and uncertainty quantification. A natural extension is the integration of COREL’s conformal prediction framework with FIBMAP’s reconstruction pipeline, providing calibrated, distribution-free prediction intervals for imputation maps. As discussed in [Sec. 6.5](#), extending conformal prediction from the forecasting to the imputation setting requires synthesising entry-wise conformal imputation frameworks [Gui, R. Barber, and Ma, [2023](#); Ke et al., [2025](#)] with the dependence-aware methodology that COREL instantiates. Specific challenges include defining conformity scores for imputed entries where the masking structure couples the locations of missing data to the error distribution, and establishing whether coverage can be conditioned on the missingness type and patterns. In the context of cardiac mapping, such an extension would enable clinicians to rigorously assess the reliability of reconstructed activation patterns, identifying regions of high predictive confidence that could support more targeted ablation strategies while flagging unreliable reconstructions to avoid misguided treatment decisions.

Integration of reconstruction and graph topology learning. FIBMAP operates on a static anatomical graph, yet the functional connectivity governing electrical propagation evolves as the arrhythmia progresses. Integrating dynamic graph learning directly within the reconstruction model would enable joint estimation of the spatiotemporal signal and the underlying topology, allowing conduction pathways to adapt during inference. This could draw on the learnable sparse dependency graphs developed in COREL or the adaptive online estimation framework of ADACGP, providing a more interpretable representation of fibrillation mechanisms than fixed anatomical connectivity. As discussed in [Sec. 7.4](#), ADACGP itself admits several extensions, including non-linear adaptive filtering to better capture chaotic fibrillation dynamics and domain-specific constraints on the graph shift operator to align learned topologies with established biophysical models of cardiac conduction.

Diffusion models for spatiotemporal reconstruction. While conformal prediction provides post-hoc uncertainty quantification for any point predictor, generative models offer an alternative in which distributional estimates emerge natively from the reconstruction process. The diffusion-based generative models demonstrated in [Chapter 4](#) for temporal imputation may offer substantial advantages in the spatiotemporal reconstruction setting. Recent work has increasingly explored the integration of denoising diffusion probabilistic models with graph-based spatiotemporal architectures for imputation tasks [M. Liu et al., [2023](#); Hu et al., [2024](#); Yakun Chen et al., [2024](#); Z. Liu, X. Zhao, and You Song, [2025](#)].

Extending FIBMAP to operate as a conditional diffusion model is promising for two reasons: first, diffusion models have demonstrated high-fidelity reconstructions in related domains [L. Chen et

al., 2023; Martinů and Šimánek, 2024; Yang et al., 2023]; and second, their ability to draw multiple samples from the learned conditional distribution enables exploration of the space of plausible reconstruction solutions [Jenkins, Z. Chen, et al., 2023; Yang et al., 2023]. In the context of cardiac fibrillation mapping, sampling alternative reconstructions consistent with the observed sparse measurements could reveal variations in wavefront organisation, offering complementary insights into the mechanisms sustaining the arrhythmia beyond those available from a single point reconstruction. While current diffusion-based approaches remain computationally expensive, recent advances in flow matching [Lipman et al., 2022], rectified flows [X. Liu, Gong, and Q. Liu, 2023], and consistency models [Yang Song et al., 2023] offer pathways to substantially reduce sampling costs through straighter transport paths and few-step or single-step generation, making generative spatiotemporal reconstruction increasingly viable for clinical applications.

Active learning for guided catheter placement. A related but distinct challenge not addressed in this thesis is optimising the *acquisition* of sparse measurements during mapping procedures. At test time, FIBMAP adapts to individual patients through learned embedding parameters, but the sequence and spatial distribution of catheter placements was not informed by the model’s informational needs. Guiding the catheter to locations that maximally reduce reconstruction uncertainty could enable faster convergence of patient-specific embeddings and more accurate global reconstructions from fewer measurements. Advances in active learning [Settles, 2009; Ren et al., 2021], particularly approaches that balance exploration and exploitation over spatial domains, may prove a fruitful avenue for developing such acquisition strategies.

Graph partitioning for *in silico* ablation planning. A particularly exciting direction enabled by this pipeline is the use of graph partitioning techniques to simulate and optimise ablation procedures on reconstructed atrial dynamics. If FIBMAP provides a global spatiotemporal reconstruction and ADACGP identifies the dynamic functional connectivity sustaining the arrhythmia, ablation can be modelled as the removal of edges or nodes from the learned graph, and the optimal lesion set becomes a combinatorial partitioning problem over that graph.

This formulation connects directly to the author’s work on constrained graph partitioning for student group formation [Jenkins, Rainbow, et al., 2026], where a variable neighbourhood search heuristic was developed that achieves near-optimal solutions to NP-hard partitioning objectives under multiple simultaneous constraints. This algorithm has been deployed over the past three years in the Department of Electrical and Electronic Engineering at Imperial College London to form skill-diverse and demographically fair tutorial groups from routinely collected academic data.

Though the domains differ substantially, the same algorithmic framework (constrained combinatorial optimisation over graph structures with domain-specific objectives) could be repurposed for ablation planning, where the objective shifts from maximising intra-group skill diversity to identifying the minimal set of edges or nodes whose removal would restore an orderly periodic rhythm to the chamber. Realising this vision requires two components: suitable optimisation metrics and efficient simulation strategies. Combined with the dynamic connectivity estimates from ADACGP to identify temporally stable propagation pathways as candidate ablation targets, this approach would enable clinicians to evaluate personalised ablation strategies *in silico* before committing to irreversible tissue destruction, an intra-procedural alternative to digital twin paradigms [Boyle et al., 2019].

8.2.2 Clinical Translation

Prospective clinical validation. The translational showcase demonstrated that FIBMAP can reconstruct personalised global dynamics from routine contact mapping recordings, establishing clinical feasibility. The natural next step is prospective validation with contemporaneous contact mapping recordings and a suitable global ground truth, enabling direct temporal validation across diverse patient phenotypes and AF organisation levels. Ultimately, whether imputation mapping translates into improved ablation success rates must be determined through prospective clinical trials.

Real-time intra-procedural deployment. For imputation mapping to be integrated into clinical workflows, the reconstruction pipeline must address both computational and representational requirements compatible with real-time use. A practical prerequisite is the availability of the patient-specific graph structure on which FIBMAP operates. As discussed in Fig. 3.5.1, electroanatomical mapping platforms construct the endocardial surface mesh progressively as the catheter is swept across the chamber, with each contact point contributing both a spatial coordinate and a local electrogram. Since the graph is not fully available at the start of the procedure, the current approach may restrict FIBMAP to post-hoc reconstruction once a sufficiently complete geometry has been acquired, rather than truly real-time use.

One potential avenue to address this is initialising the graph from a coarse sweep of the geometry performed at the start of the procedure, before detailed electrogram acquisition begins, and refining the mesh as additional contact points are registered. Alternatively, integration with pre-procedural CT or MRI surface models through image registration [Sy, Thiagalingam, and Stiles, 2012; Narayan and John, 2024] could provide a complete graph from the outset, though this would

reintroduce a dependency on pre-procedural imaging that the current FIBMAP framework avoids. With a graph established, FIBMAP’s recurrent architecture processes each time step sequentially, with the computational cost per step dominated by the message-passing operations.

While the current implementation is sufficiently efficient for retrospective analysis, optimisation for real-time deployment will require attention to hardware acceleration, model compression, and the development of streaming inference strategies that update reconstructions incrementally as new catheter measurements are acquired. A complementary direction is online fine-tuning of the patient-specific embedding parameters during the procedure itself, extending the hybrid global-local adaptation strategy to a streaming setting where embeddings are updated continuously as new observations arrive, analogous in spirit to the constant per-iteration adaptive updates developed for ADACGP (Chapter 7).

Integration with clinical annotation and decision support. By providing comprehensive spatiotemporal data previously inaccessible during clinical procedures, imputation maps enable visual identification of driver regions and may enhance automated annotation algorithms that benefit from the complete spatial context [Liao et al., 2021; Seitz et al., 2022; Manongi, J. Kim, and Goldbarg, 2023]. Integrating FIBMAP with existing electrogram annotation tools could enable robust electrophenotyping based on unique patient AF dynamics [F. S. Ng et al., 2020], facilitating personalised ablation strategies. Furthermore, combining reconstructed dynamics with uncertainty quantification from COREL and connectivity characterisation from ADACGP could inform a decision support system that presents clinicians with a comprehensive view of how wavefronts propagate, the confidence of the reconstruction, and how stable the identified conduction pathways are.

Multi-arrhythmia generalisation. The methods developed in this thesis have been validated on AF and VF, but the graph-based spatiotemporal framework is in principle applicable to any cardiac arrhythmia where spatial propagation patterns are of clinical interest. Extending FIBMAP to atrial tachycardia, ventricular tachycardia, and other complex arrhythmias would broaden the clinical utility of imputation mapping and may accelerate existing mapping workflows for these arrhythmias. Even for organised arrhythmias where sequential contact mapping can produce reliable activation maps, the process of acquiring sufficiently dense catheter coverage to construct high-resolution local activation time maps remains time-consuming [Özgül et al., 2023; Narayan and John, 2024]. By reconstructing global activation patterns from sparser catheter placements, a FIBMAP-like architecture could reduce the number of sequential acquisitions required, shortening procedure times while maintaining or improving map fidelity.

8.3 Closing Remarks

This thesis set out to advance graph-based spatiotemporal machine learning and demonstrate its application to the long-standing clinical challenge of cardiac fibrillation mapping. The methods developed herein, FIBMAP for reconstruction, COREL for calibrated uncertainty quantification, and ADACGP for dynamic topology learning, form a progressive pipeline that addresses complementary aspects of this challenge. Each contribution also advances the broader methodological landscape: FIBMAP demonstrates that time-and-space architectures with hybrid global-local parameterisation can solve large-scale spatiotemporal imputation problems on irregular domains; COREL establishes that relational structure in prediction residuals is a valuable and underexploited source of information for conformal prediction; and ADACGP introduces the first online algorithm capable of achieving true sparsity in adaptive graph topology estimation. Within cardiac electrophysiology, the clinical utility is already concrete: FIBMAP reconstructs global AF dynamics from sparse routine catheter recordings, a capability that did not previously exist and for which patent applications have been filed, while ADACGP tracks dynamic functional connectivity from fibrillation recordings where batch methods fail. Together, these results establish that graph-based spatiotemporal machine learning offers a principled and effective framework for recovering, analysing, and characterising complex dynamics on irregular spatial domains. Prospective clinical validation and full pipeline integration represent the natural next steps, as these methods progress from validated components toward integration into the electrophysiology laboratory.

Bibliography

- Ahmad, Tabia and Panagiotis N Papadopoulos [2022]. “Prediction of cascading failures and simultaneous learning of functional connectivity in power system”. In: *Proceedings of the IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*. IEEE, pp. 1–5.
- Alcaine, Alejandro, Michela Mase, et al. [2016]. “A multi-variate predictability framework to assess invasive cardiac activity and interactions during atrial fibrillation”. In: *IEEE Transactions on Biomedical Engineering* 64.5, pp. 1157–1168.
- Alcaraz, Juan Lopez and Nils Strodthoff [2023]. “Diffusion-based time series imputation and forecasting with structured state space models”. In: *Transactions on Machine Learning Research*.
- Allessie, Maurits A et al. [2010]. “Electropathological substrate of long-standing persistent atrial fibrillation in patients with structural heart disease: longitudinal dissociation”. In: *Circulation: Arrhythmia and Electrophysiology* 3.6, pp. 606–615.
- Alon, Uri and Eran Yahav [2021]. “On the bottleneck of graph neural networks and its practical implications”. In: *Proceedings of The International Conference on Learning Representations*.
- Alteriis, Giuseppe de, Oliver Sherwood, et al. [2025]. “DySCo: A general framework for dynamic functional connectivity”. In: *PLoS Computational Biology* 21.3, e1012795.
- Angelopoulos, Anastasios N, Rina Foygel Barber, and Stephen Bates [2024]. “Theoretical foundations of conformal prediction”. In: *arXiv preprint arXiv:2411.11824*.
- Angelopoulos, Anastasios N, Stephen Bates, et al. [2023]. “Conformal prediction: a gentle introduction”. In: *Foundations and Trends® in Machine Learning* 16.4, pp. 494–591.
- Atluri, Gowtham, Anuj Karpatne, and Vipin Kumar [Aug. 2018]. “Spatio-temporal data mining: a survey of problems and methods”. In: *ACM Computing Surveys* 51.4, pp. 1–41.
- Atwood, James and Don Towsley [2016]. “Diffusion-convolutional neural networks”. In: *Proceedings of the International Conference on Neural Information Processing Systems* 29.
- Auer, Andreas et al. [2023]. “Conformal prediction for time series with modern Hopfield networks”. In: *Proceedings of the International Conference on Neural Information Processing Systems*.
- Bacciu, Davide et al. [2020]. “A gentle introduction to deep learning for graphs”. In: *Neural Networks* 129, pp. 203–221.
- Bai, Shaojie, J. Zico Kolter, and Vladlen Koltun [2018]. *An empirical evaluation of generic convolutional and recurrent networks for sequence modeling*. arXiv: [1803.01271](https://arxiv.org/abs/1803.01271) [cs.LG].
- Baingana, Brian and Georgios B Giannakis [2016]. “Tracking switched dynamic network topologies from information cascades”. In: *IEEE Transactions on Signal Processing* 65.4, pp. 985–997.
- Bansal, Parikshit, Prathamesh Deshpande, and Sunita Sarawagi [2021]. “Missing value imputation on multidimensional time series”. In: *Proceedings of the Very Large Data Bases (VLDB) Endowment* 14, pp. 2533–2545.

- Barabási, Albert-László and Réka Albert [1999]. “Emergence of scaling in random networks”. In: *Science* 286.5439, pp. 509–512.
- Barber, Rina Foygel et al. [2023]. “Conformal prediction beyond exchangeability”. In: *The Annals of Statistics* 51.2, pp. 816–845.
- Battaglia, Peter W et al. [2018]. “Relational inductive biases, deep learning, and graph networks”. In: *arXiv preprint arXiv:1806.01261*.
- Belkin, Mikhail and Partha Niyogi [2003]. “Laplacian eigenmaps for dimensionality reduction and data representation”. In: *Neural Computation* 15.6, pp. 1373–1396.
- Bellman, Richard E. [1961]. *Adaptive Control Processes: A Guided Tour*. Princeton University Press. ISBN: 9781400874668.
- Bengio, Yoshua, Nicholas Léonard, and Aaron Courville [2013]. “Estimating or propagating gradients through stochastic neurons for conditional computation”. In: *arXiv preprint arXiv:1308.3432*.
- Bengio, Yoshua, Patrice Simard, and Paolo Frasconi [1994]. “Learning long-term dependencies with gradient descent is difficult”. In: *IEEE Transactions on Neural Networks* 5.2, pp. 157–166.
- Benidis, Konstantinos et al. [2022]. “Deep learning for time series forecasting: tutorial and literature survey”. In: *ACM Computing Surveys* 55.6.
- Bernardini, Fausto et al. [1999]. “The ball-pivoting algorithm for surface reconstruction”. In: *IEEE Transactions on Visualization and Computer Graphics* 5.4, pp. 349–359.
- Besl, Paul J and Neil D McKay [1992]. “Method for registration of 3-D shapes”. In: *Proceedings of Sensor fusion IV: Control Paradigms and Data Structures* 1611, pp. 586–606.
- Bézenac, Emmanuel de et al. [2020]. “Normalizing Kalman filters for multivariate time series analysis”. In: *Proceedings of the International Conference on Neural Information Processing Systems* 33, pp. 2995–3007.
- Bottou, Léon, Frank E. Curtis, and Jorge Nocedal [2018]. “Optimization methods for large-scale machine learning”. In: *SIAM Review* 60.2, pp. 223–311.
- Boukis, Christos, Danilo P Mandic, et al. [2010]. “A modified Armijo rule for the online selection of learning rate of the LMS algorithm”. In: *Digital Signal Processing* 20.3, pp. 630–639.
- Boyle, Patrick M et al. [2019]. “Computationally guided personalized targeted ablation of persistent atrial fibrillation”. In: *Nature Biomedical Engineering* 3.11, pp. 870–879.
- Brachmann, JHJW et al. [2019]. “Prospective randomized comparison of rotor ablation vs conventional ablation for treatment of persistent atrial fibrillation—The REAFFIRM trial”. In: *Heart Rhythm* 16.6, pp. 963–965.
- Bray, Mark-Anthony et al. [2001]. “Experimental and theoretical analysis of phase singularity dynamics in cardiac tissue”. In: *Journal of Cardiovascular Electrophysiology* 12.6, pp. 716–722.
- Bresson, Xavier and Thomas Laurent [2017]. “Residual gated graph convnets”. In: *arXiv preprint arXiv:1711.07553*.
- Bronstein, Michael M, Joan Bruna, Taco Cohen, et al. [2021]. “Geometric deep learning: grids, groups, graphs, geodesics, and gauges”. In: *arXiv preprint arXiv:2104.13478*.
- Bronstein, Michael M, Joan Bruna, Yann LeCun, et al. [2017]. “Geometric deep learning: going beyond euclidean data”. In: *IEEE Signal Processing Magazine* 34.4, pp. 18–42.

- Bruna, Joan et al. [2014]. “Spectral networks and deep locally connected networks on graphs”. In: *Proceedings of The International Conference on Learning Representations*.
- Buch, Eric et al. [2016]. “Long-term clinical outcomes of focal impulse and rotor modulation for treatment of atrial fibrillation: a multicenter experience”. In: *Heart Rhythm* 13.3, pp. 636–641.
- Cai, Deng et al. [2010]. “Graph regularized nonnegative matrix factorization for data representation”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33.8, pp. 1548–1560.
- Cao, Wei et al. [2018]. “BRITS: Bidirectional recurrent imputation for time series”. In: *Proceedings of the International Conference on Neural Information Processing Systems*. Vol. 31.
- Cardoso, Rhanderson et al. [2022]. “Catheter ablation is superior to antiarrhythmic drugs as first-line treatment for atrial fibrillation: a systematic review and meta-analysis”. In: *Arquivos Brasileiros de Cardiologia* 119, pp. 87–94.
- Chamberlain, Ben et al. [July 2021]. “GRAND: Graph neural diffusion”. In: *Proceedings of the International Conference on Machine Learning*, pp. 1407–1418.
- Chen, Lei et al. [2023]. “SwinRDM: Integrate SwinRNN with diffusion model towards high-resolution and high-quality weather forecasting”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 37. 1, pp. 322–330.
- Chen, Sheng, Stephen A Billings, and PM Grant [1990]. “Non-linear system identification using neural networks”. In: *International Journal of Control* 51.6, pp. 1191–1214.
- Chen, Yakun et al. [2024]. “A temporally disentangled contrastive diffusion model for spatiotemporal imputation”. In: *CAAI Transactions on Intelligence Technology*.
- Chen, Yilun, Yuantao Gu, and Alfred O Hero [2009]. “Sparse LMS for system identification”. In: *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, pp. 3125–3128.
- Chen, Yilun, Ami Wiesel, and Alfred O. Hero [2011]. “Robust shrinkage estimation of high-dimensional covariance matrices”. In: *IEEE Transactions on Signal Processing* 59.9, pp. 4097–4107.
- Chen, Zehua, Bruno Scalzo Dees, and Danilo P. Mandic [2020]. “A probabilistic beat-to-beat filtering model for continuous and accurate blood pressure estimation”. In: *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, pp. 1–6.
- Chen, Zehua, Yihan Wu, et al. [2022]. “ResGrad: Residual denoising diffusion probabilistic models for text to speech”. In: *arXiv preprint arXiv:2212.14518*.
- Cheniti, Ghassen et al. [2018]. “Atrial fibrillation mechanisms and implications for catheter ablation”. In: *Frontiers in Physiology* 9, p. 1458.
- Chiang, Wei-Lin et al. [2019]. “Cluster-GCN: An efficient algorithm for training deep and large graph convolutional networks”. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 257–266.
- Cho, Kyunghyun, Bart Van Merriënboer, et al. [2014]. “On the properties of neural machine translation: encoder-decoder approaches”. In: *arXiv preprint arXiv:1409.1259*.

- Cho, Kyunghyun, Bart Van Merriënboer, et al. [2014]. “Learning phrase representations using RNN encoder–decoder for statistical machine translation”. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1724–1734.
- Christov, Ivaylo I [2004]. “Real time electrocardiogram QRS detection using combined adaptive threshold”. In: *BioMedical Engineering OnLine* 3.1.
- Chung, Fan RK [1997]. *Spectral graph theory*. Vol. 92. American Mathematical Society.
- Cini, Andrea, Alexander Jenkins, et al. [2025]. “Relational conformal prediction for correlated time series”. In: *Proceedings of the International Conference on Machine Learning*.
- Cini, Andrea and Ivan Marisca [2022]. *Torch Spatiotemporal*. URL: <https://github.com/TorchSpatiotemporal/tsl>.
- Cini, Andrea, Ivan Marisca, and Cesare Alippi [2022]. “Filling the g_ap_s: multivariate time series imputation by graph neural networks”. In: *Proceedings of the International Conference on Learning Representations*.
- Cini, Andrea, Ivan Marisca, Filippo Maria Bianchi, et al. [2023]. “Scalable spatiotemporal graph neural networks”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 37. Chap. 6, pp. 7218–7226.
- Cini, Andrea, Ivan Marisca, Daniele Zambon, et al. [2024]. “Taming local effects in graph-based spatiotemporal forecasting”. In: *Proceedings of the International Conference on Neural Information Processing Systems* 36.
- [2025]. “Graph deep learning for time series forecasting”. In: *ACM Computing Surveys* 57.12, pp. 1–34.
- Cini, Andrea, Daniele Zambon, and Cesare Alippi [2023]. “Sparse graph learning from spatiotemporal time series”. In: *Journal of Machine Learning Research* 24.242, pp. 1–36.
- Clarkson, Jase [2023]. “Distribution free prediction sets for node classification”. In: *Proceedings of the International Conference on Machine Learning*, pp. 6268–6278.
- Commission for Energy Regulation [2016]. “CER smart metering project - electricity customer behaviour trial, 2009-2010 [dataset]”. In: *Irish Social Science Data Archive*. SN: 0012-00. URL: <https://www.ucd.ie/issda/data/commissionforenergyregulationcer>.
- Connor, Jerome T, R Douglas Martin, and Les E Atlas [1994]. “Recurrent neural networks and robust time series prediction”. In: *IEEE Transactions on Neural Networks* 5.2, pp. 240–254.
- Coumel, Philippe [1994]. “Paroxysmal atrial fibrillation: a disorder of autonomic tone?” In: *European Heart Journal* 15.suppl_A, pp. 9–16.
- Dabney, Will et al. [2018]. “Implicit quantile networks for distributional reinforcement learning”. In: *Proceedings of the International Conference on Machine Learning*, pp. 1096–1105.
- Davidenko, Jorge M et al. [1992]. “Stationary and drifting spiral waves of excitation in isolated cardiac muscle”. In: *Nature* 355.6358, pp. 349–351.
- Davies, Harry J et al. [2024]. “The deep-match framework: R-peak detection in ear-ECG”. In: *IEEE Transactions on Biomedical Engineering* 71.7, pp. 2014–2021.
- Davies, Harry J. et al. [2022]. “Wearable in-ear PPG: detailed respiratory variations enable classification of COPD”. In: *IEEE Transactions on Biomedical Engineering* 69.7, pp. 2390–2400.

- De Groot, Natasja MS et al. [2022]. “Critical appraisal of technologies to assess electrical activity during atrial fibrillation: a position paper from the European heart rhythm association and European society of cardiology working group on eCardiology in collaboration with the heart rhythm society, asia pacific heart rhythm society, Latin American heart rhythm society and computing in cardiology”. In: *EP Europace* 24.2, pp. 313–330.
- Deb, Brototo et al. [2021]. “Identifying atrial fibrillation mechanisms for personalized medicine”. In: *Journal of Clinical Medicine* 10.23, p. 5679.
- Dees, Bruno Scalzo et al. [2019]. “Unitary shift operators on a graph”. In: *arXiv preprint arXiv:1909.05767*.
- Defferrard, Michaël, Xavier Bresson, and Pierre Vandergheynst [2016]. “Convolutional neural networks on graphs with fast localized spectral filtering”. In: *Proceedings of the International Conference on Neural Information Processing Systems* 29.
- Di Lorenzo, Paolo, Paolo Banelli, et al. [2017]. “Distributed adaptive learning of graph signals”. In: *IEEE Transactions on Signal Processing* 65.16, pp. 4193–4208.
- Di Lorenzo, Paolo, Sergio Barbarossa, et al. [2016]. “Adaptive least mean squares estimation of graph signals”. In: *IEEE Transactions on Signal and Information Processing over Networks* 2.4, pp. 555–568.
- Dong, Xiaowen, Dorina Thanou, et al. [2016]. “Learning Laplacian matrix in smooth graph signal representations”. In: *IEEE Transactions on Signal Processing* 64.23, pp. 6160–6173.
- Dong, Xinzhen et al. [2019]. “An improved method of handling missing values in the analysis of sample entropy for continuous monitoring of physiological signals”. In: *Entropy* 21, p. 274.
- Donoho, D.L. [1995]. “De-noising by soft-thresholding”. In: *IEEE Transactions on Information Theory* 41.3, pp. 613–627.
- Du, Wenjie, David Côté, and Yan Liu [2023]. “SAITS: Self-attention-based imputation for time series”. In: *Expert Systems with Applications* 219, p. 119619.
- Dwivedi, Vijay Prakash and Xavier Bresson [2021]. “A generalization of transformer networks to graphs”. In: *Proceedings of the AAAI 2021 Workshop on Deep Learning on Graphs: Methods and Applications (DLG-AAAI 2021)*.
- Eckstein, Jens et al. [2013]. “Transmural conduction is the predominant mechanism of breakthrough during atrial fibrillation: evidence from simultaneous endo-epicardial high-density activation mapping”. In: *Circulation: Arrhythmia and Electrophysiology* 6.2, pp. 334–341.
- Efimov, Igor R, Vladimir P Nikolski, and Guy Salama [2004]. “Optical imaging of the heart”. In: *Circulation Research* 95.1, pp. 21–33.
- Elman, Jeffrey L [1990]. “Finding structure in time”. In: *Cognitive Science* 14.2, pp. 179–211.
- Erdos, Paul, Alfréd Rényi, et al. [1960]. “On the evolution of random graphs”. In: *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* 5.1, pp. 17–60.
- Everett IV, Thomas H et al. [2001]. “Assessment of global atrial fibrillation organization to optimize timing of atrial defibrillation”. In: *Circulation* 103.23, pp. 2857–2861.
- Falcon, William and The PyTorch Lightning team [Mar. 2019]. *PyTorch Lightning*. Version 1.4. URL: <https://github.com/PyTorchLightning/pytorch-lightning>.

- Feldman, Shai, Stephen Bates, and Yaniv Romano [2023]. “Calibrated multiple-output quantile regression with representation learning”. In: *Journal of Machine Learning Research* 24.24, pp. 1–48.
- Fenton, Flavio and Alain Karma [Mar. 1998]. “Vortex dynamics in three-dimensional continuous myocardium with fiber rotation: filament instability and fibrillation”. In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 8.1, pp. 20–47. ISSN: 1054-1500.
- Fenton, Flavio H et al. [2002]. “Multiple mechanisms of spiral wave breakup in a model of cardiac electrical activity”. In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 12.3, pp. 852–892.
- Fey, Matthias and Jan Eric Lenssen [2019]. “Fast graph representation learning with PyTorch Geometric”. In: *arXiv preprint arXiv:1903.02428*.
- Filgueiras-Rama, David and José Jalife [2016]. “Structural and functional bases of cardiac fibrillation: differences and similarities between atria and ventricles”. In: *JACC: Clinical Electrophysiology* 2.1, pp. 1–13.
- Forni, Mario et al. [2000]. “The generalized dynamic-factor model: identification and estimation”. In: *Review of Economics and Statistics* 82.4, pp. 540–554.
- Frasconi, Paolo, Marco Gori, and Alessandro Sperduti [1998]. “A general framework for adaptive processing of data structures”. In: *IEEE Transactions on Neural Networks* 9.5, pp. 768–786.
- Friedman, Jerome, Trevor Hastie, and Robert Tibshirani [Dec. 2007]. “Sparse inverse covariance estimation with the graphical LASSO”. In: *Biostatistics* 9.3, pp. 432–441. ISSN: 1465-4644.
- Gafni, Eli M and Dimitri P Bertsekas [1984]. “Two-metric projection methods for constrained optimization”. In: *SIAM Journal on Control and Optimization* 22.6, pp. 936–964.
- Gao, Jianfei and Bruno Ribeiro [2022]. “On the equivalence between temporal and static equivariant graph representations”. In: *Proceedings of the International Conference on Machine Learning*. PMLR, pp. 7052–7076.
- Gasthaus, Jan et al. [2019]. “Probabilistic forecasting with spline quantile function RNNs”. In: *Proceedings of the International Conference on Artificial Intelligence and Statistics*, pp. 1901–1910.
- Geman, Stuart, Elie Bienenstock, and René Doursat [1992]. “Neural networks and the bias/variance dilemma”. In: *Neural Computation* 4.1, pp. 1–58.
- Gibbs, Isaac and Emmanuel Candes [2021]. “Adaptive conformal inference under distribution shift”. In: *Proceedings of the International Conference on Neural Information Processing Systems* 34, pp. 1660–1672.
- Gillis, Anne M et al. [2013]. “Management of atrial fibrillation in the year 2033: new concepts, tools, and applications leading to personalized medicine”. In: *Canadian Journal of Cardiology* 29.10, pp. 1141–1146.
- Gilmer, Justin et al. [2017]. “Neural message passing for quantum chemistry”. In: *Proceedings of the International Conference on Machine Learning*, pp. 1263–1272.
- Girault, Benjamin, Paulo Gonçalves, and Éric Fleury [2015]. “Translation on graphs: an isometric shift operator”. In: *IEEE Signal Processing Letters* 22.12, pp. 2416–2420.
- Goel, Karan et al. [2022]. “It’s raw! Audio generation with state-space models”. In: *Proceedings of the International Conference on Machine Learning*. Vol. 162, pp. 7616–7633.

- Goller, Christoph and Andreas Kuchler [1996]. “Learning task-dependent distributed representations by backpropagation through structure”. In: *Proceedings of International Conference on Neural Networks (ICNN’96)*. Vol. 1. IEEE, pp. 347–352.
- Gori, Marco, Gabriele Monfardini, and Franco Scarselli [2005]. “A new model for learning in graph domains”. In: *Proceedings of the IEEE International Joint Conference on Neural Networks*. Vol. 2. IEEE, pp. 729–734.
- Gouttes, Adèle et al. [2021]. “Probabilistic time series forecasting with implicit quantile networks”. In: *arXiv preprint arXiv:2107.03743*.
- Grace, Andrew et al. [2019]. “High-resolution noncontact charge-density mapping of endocardial activation”. In: *JCI Insight* 4.6.
- Granger, Clive WJ [1969]. “Investigating causal relations by econometric models and cross-spectral methods”. In: *Econometrica: Journal of the Econometric Society*, pp. 424–438.
- Grassi, Francesco et al. [2018]. “A time-vertex signal processing framework: scalable processing and meaningful representations for time-series on graphs”. In: *IEEE Transactions on Signal Processing* 66.3, pp. 817–829.
- Gravina, Alessio et al. [2024]. “Long range propagation on continuous-time dynamic graphs”. In: *Proceedings of the International Conference on Machine Learning*.
- Gray, Richard A, Arkady M Pertsov, and José Jalife [1998]. “Spatial and temporal organization during cardiac fibrillation”. In: *Nature* 392.6671, pp. 75–78.
- Gray, Richard A. and Pras Pathmanathan [Mar. 2018]. “Patient-specific cardiovascular computational modeling: diversity of personalization and challenges”. In: *Journal of Cardiovascular Translational Research* 11.2, pp. 80–88.
- Gu, Albert, Karan Goel, and Christopher Re [2022]. “Efficiently modeling long sequences with structured state spaces”. In: *Proceedings of the International Conference on Learning Representations*.
- Gui, Yu, Rina Barber, and Cong Ma [2023]. “Conformalized matrix completion”. In: *Proceedings of the International Conference on Neural Information Processing Systems* 36, pp. 4820–4844.
- Haissaguerre, Michel et al. [Sept. 1998]. “Spontaneous initiation of atrial fibrillation by ectopic beats originating in the pulmonary veins”. In: *New England Journal of Medicine* 339.10, pp. 659–666.
- Hamdi, Ali et al. [Apr. 2021]. “Spatiotemporal data mining: a survey on challenges and open problems”. In: *Artificial Intelligence Review* 55.2, pp. 1441–1488.
- Hamilton, Will, Zhitao Ying, and Jure Leskovec [2017]. “Inductive representation learning on large graphs”. In: *Proceedings of the International Conference on Neural Information Processing Systems* 30.
- Handa, Balvinder S, Xinyang Li, et al. [2021]. “Ventricular fibrillation mechanism and global fibrillatory organization are determined by gap junction coupling and fibrosis pattern”. In: *Cardiovascular Research* 117.4, pp. 1078–1090.
- Handa, Balvinder S, Xinyang Li, et al. [2020]. “Granger causality-based analysis for classification of fibrillation mechanisms and localization of rotational drivers”. In: *Circulation: Arrhythmia and Electrophysiology* 13.3, e008237.

- Harris, Charles R et al. [Sept. 2020]. “Array programming with NumPy”. In: *Nature* 585.7825, pp. 357–362.
- Hasan, Asif, Digvijay Pandey, and Azizuddin Khan [2021]. “Application of EEG time-varying networks in the evaluation of dynamic functional brain networks”. In: *Augmented Human Research* 6, pp. 1–8.
- Hayes, Monson H [1996]. *Statistical digital signal processing and modeling*. John Wiley & Sons.
- Haykin, Simon S [2008]. *Adaptive filter theory*. Pearson Education India.
- He, Wenchong and Zhe Jiang [2023]. “Uncertainty quantification of deep learning for spatiotemporal data: challenges and opportunities”. In: *Proceedings of the 2nd KDD Workshop on Uncertainty Reasoning and Quantification in Decision Making (UDM-KDD’23)*.
- Heijman, Jordi et al. [2014]. “Cellular and molecular electrophysiology of atrial fibrillation initiation, maintenance, and progression”. In: *Circulation Research* 114.9, pp. 1483–1499.
- Hindricks, Gerhard et al. [2021]. “2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS) The Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) Developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC”. In: *European Heart Journal* 42.5, pp. 373–498.
- Ho, Jonathan and Stefano Ermon [2016]. “Generative adversarial imitation learning”. In: *Proceedings of the International Conference on Neural Information Processing Systems*. Vol. 29.
- Ho, Jonathan, Ajay Jain, and Pieter Abbeel [2020]. “Denoising diffusion probabilistic models”. In: *Proceedings of the International Conference on Neural Information Processing Systems*. Vol. 33, pp. 6840–6851.
- Ho, Jonathan and Tim Salimans [2022]. “Classifier-free diffusion guidance”. In: *arXiv preprint arXiv:2207.12598*.
- Ho, Siew Yen, Robert H Anderson, and Damián Sánchez-Quintana [2002]. “Atrial structure and fibres: morphologic bases of atrial conduction”. In: *Cardiovascular Research* 54.2, pp. 325–336.
- Hochreiter, Sepp and Jürgen Schmidhuber [1997]. “Long short-term memory”. In: *Neural Computation* 9.8, pp. 1735–1780.
- Hong, Kathryn Lauren et al. [2019]. “High-resolution mapping of the atria using the HD Grid catheter”. In: *HeartRhythm Case Reports* 5.7, pp. 351–353.
- Hsieh, Yu-Cheng, Jiunn-Cherng Lin, et al. [2016]. “Gap junction modifier rotigaptide decreases the susceptibility to ventricular arrhythmia by enhancing conduction velocity and suppressing discordant alternans during therapeutic hypothermia in isolated rabbit hearts”. In: *Heart Rhythm* 13.1, pp. 251–261.
- Hu, Junfeng et al. [2024]. “Towards unifying diffusion models for probabilistic spatio-temporal graph learning”. In: *Proceedings of the 32nd ACM International Conference on Advances in Geographic Information Systems*, pp. 135–146.
- Huang, Kexin et al. [2024]. “Uncertainty quantification over graph with conformalized graph neural networks”. In: *Proceedings of the International Conference on Neural Information Processing Systems* 36.
- Huang, Yu et al. [2021]. “What makes multi-modal learning better than single (provably)”. In: *Proceedings of the International Conference on Neural Information Processing Systems* 34, pp. 10944–10956.
- Hyndman, Rob J and George Athanasopoulos [2018]. *Forecasting: principles and practice*. OTexts.

- Isufi, Elvin et al. [2019]. “Forecasting time series with VARMA recursions on graphs”. In: *IEEE Transactions on Signal Processing* 67.18, pp. 4870–4885.
- Jadidi, Amir S. et al. [2016]. “Ablation of persistent atrial fibrillation targeting low-voltage areas with selective activation characteristics”. In: *Circulation: Arrhythmia and Electrophysiology* 9.3, e002962.
- Jais, Pierre et al. [2004]. “Technique and results of linear ablation at the mitral isthmus”. In: *Circulation* 110.19, pp. 2996–3002.
- Janse, MJ [1997]. “Why does atrial fibrillation occur?” In: *European Heart Journal* 18.suppl_C, pp. 12–18.
- Jenkins, Alexander, Zehua Chen, et al. [2023]. “Improving diffusion models for ECG imputation with an augmented template prior”. In: *Proceedings of the ICML Workshop on Interpretable Machine Learning in Healthcare*.
- Jenkins, Alexander, Andrea Cini, Joseph Barker, Danilo Mandic, et al. [2025]. “Mapping the dynamics of atrial fibrillation with spatiotemporal graph neural networks”. In: *Proceedings of the NeurIPS Workshop on Learning from Time Series for Health*.
- Jenkins, Alexander, Andrea Cini, Joseph Barker, Alexander Sharp, et al. [2025]. “Learning to map global atrial fibrillation dynamics from sparse measurements”. In: *EP Europace*. Under Review.
- Jenkins, Alexander, Emma Rainbow, et al. [2026]. “A graph-theoretic approach to fair and skill-diverse student group formation”. In: *IEEE Signal Processing Magazine*.
- Jenkins, Alexander, Thiernithi Varidhisai, et al. [2025]. “Online graph topology learning via time-vertex adaptive filters: from theory to cardiac fibrillation”. In: *IEEE Transactions on Signal and Information Processing over Networks* 11, pp. 965–979.
- Jiang, Hanyang, Yao Xie, and Feng Qiu [2024]. “Spatio-temporal conformal prediction for power outage data”. In: *arXiv preprint arXiv:2411.17099*.
- Jin, Guangyin et al. [2024]. “Spatio-temporal graph neural networks for predictive learning in urban computing: a survey”. In: *IEEE Transactions on Knowledge and Data Engineering* 36.10, pp. 5388–5408.
- Jin, Ming et al. [2024]. “A survey on graph neural networks for time series: forecasting, classification, imputation, and anomaly detection”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 46.12, pp. 10466–10485.
- Kan, Kelvin et al. [2022]. “Multivariate quantile function forecaster”. In: *Proceedings of the International Conference on Artificial Intelligence and Statistics*, pp. 10603–10621.
- Kappadan, Vineesh et al. [2020]. “High-resolution optical measurement of cardiac restitution, contraction, and fibrillation dynamics in beating vs. blebbistatin-uncoupled isolated rabbit hearts”. In: *Frontiers in Physiology* 11, p. 464.
- Katsanos, Aristeidis H et al. [2020]. “Stroke prevention in atrial fibrillation: looking forward”. In: *Circulation* 142.24, pp. 2371–2388.
- Kazi, Anees et al. [2022]. “Differentiable graph module (DGM) for graph convolutional networks”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Ke, Yuxuan et al. [2025]. “Matrix factorization-based solar spectral irradiance missing data imputation with uncertainty quantification”. In: *arXiv preprint arXiv:2508.04074*.

- Kim, Seung-Jean, Kwangmoo Koh, et al. [2007]. “A method for large-scale ℓ_1 -regularized least squares”. In: *IEEE Journal on Selected Topics in Signal Processing* 1.4, pp. 606–617.
- King, Benjamin et al. [2017]. “Effect of spatial resolution and filtering on mapping cardiac fibrillation”. In: *Heart Rhythm* 14.4, pp. 608–615.
- Kingma, Diederik et al. [2021]. “Variational diffusion models”. In: *Proceedings of the International Conference on Neural Information Processing Systems*. Vol. 34, pp. 21696–21707.
- Kingma, Diederik P and Jimmy Ba [2015]. “Adam: A method for stochastic optimization”. In: *Proceedings of the International Conference on Learning Representations*.
- Kipf, Thomas N. and Max Welling [2017]. “Semi-supervised classification with graph convolutional networks”. In: *Proceedings of the International Conference on Learning Representations*.
- Klabunde, Richard [2011]. *Cardiovascular physiology concepts*. Lippincott Williams & Wilkins.
- Kleber, Andre G. and Yoram Rudy [Apr. 2004]. “Basic mechanisms of cardiac impulse propagation and associated arrhythmias”. In: *Physiological Reviews* 84.2, pp. 431–488.
- Koenker, Roger and Kevin F Hallock [2001]. “Quantile regression”. In: *Journal of Economic Perspectives* 15.4, pp. 143–156.
- Kool, Wouter, Herke Van Hoof, and Max Welling [2019]. “Stochastic beams and where to find them: the Gumbel-top-K trick for sampling sequences without replacement”. In: *Proceedings of the International Conference on Machine Learning*, pp. 3499–3508.
- Kuck, Karl-Heinz et al. [2016]. “Cryoballoon or radiofrequency ablation for paroxysmal atrial fibrillation”. In: *New England Journal of Medicine* 374.23, pp. 2235–2245.
- Kyriakou, C, S Thomas, et al. [2023]. “Right and left atrial scarring in patients with atrial fibrillation: patterns, prevalence, and predictors”. In: *Heart, Lung and Circulation* 32, S212–S213.
- Lam, Remi et al. [2023]. “Learning skillful medium-range global weather forecasting”. In: *Science* 382.6677, pp. 1416–1421.
- Lamare, Rodrigo C. de and Raimundo Sampaio-Neto [2014]. “Sparsity-aware adaptive algorithms based on alternating optimization and shrinkage”. In: *IEEE Signal Processing Letters* 21.2, pp. 225–229.
- Laughner, Jacob I, Fu Siong Ng, et al. [2012]. “Processing and analysis of cardiac optical mapping data obtained with potentiometric dyes”. In: *American Journal of Physiology-Heart and Circulatory Physiology* 303.7, H753–H765.
- Le, Tan Duy, Razvan Beuran, and Yasuo Tan [2018]. “Comparison of the most influential missing data imputation algorithms for healthcare”. In: *Proceedings of the International Conference on Knowledge and Systems Engineering*.
- LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton [2015]. “Deep learning”. In: *Nature* 521.7553, pp. 436–444.
- LeCun, Yann, Bernhard Boser, et al. [1989]. “Backpropagation applied to handwritten zip code recognition”. In: *Neural Computation* 1.4, pp. 541–551.
- Ledoit, Olivier and Michael Wolf [2004]. “A well-conditioned estimator for large-dimensional covariance matrices”. In: *Journal of Multivariate Analysis* 88.2, pp. 365–411.

- Lee, Daniel and H Sebastian Seung [2000]. “Algorithms for non-negative matrix factorization”. In: *Proceedings of the 13th International Conference on Neural Information Processing Systems* 13.
- Lee, Geoffrey et al. [2014]. “Epicardial wave mapping in human long-lasting persistent atrial fibrillation: transient rotational circuits, complex wavefronts, and disorganized activity”. In: *European Heart Journal* 35.2, pp. 86–97.
- Lee, In [2017]. “Big data: dimensions, evolution, impacts, and challenges”. In: *Business Horizons* 60.3, pp. 293–303.
- Lee, Jonghyeok, Chen Xu, and Yao Xie [2025]. “Kernel-based optimally weighted conformal time-series prediction”. In: *Proceedings of the International Conference on Learning Representations*.
- Lee, Jung Myung et al. [2019]. “The electrical isolation of the left atrial posterior wall in catheter ablation of persistent atrial fibrillation”. In: *Clinical Electrophysiology* 5.11, pp. 1253–1261.
- Lee, Junghwan, Chen Xu, and Yao Xie [2024]. “Conformal prediction for time series with transformer”. In: *Proceedings of the ICML 2024 Workshop on Structured Probabilistic Inference & Generative Modeling*.
- Lee, Seungyup et al. [2017]. “Characterization of foci and breakthrough sites during persistent and long-standing persistent atrial fibrillation in patients: studies using high-density (510–512 electrodes) biatrial epicardial mapping”. In: *Journal of the American Heart Association* 6.3, e005274.
- Levie, Ron et al. [2021]. “Transferability of spectral graph convolutional neural networks”. In: *Journal of Machine Learning Research* 22.272, pp. 1–59.
- Li, Huayu et al. [2022]. “DeScoD-ECG: Deep score-based diffusion model for ECG baseline wander and noise removal”. In: *arXiv preprint arXiv:2208.00542*.
- Li, Qimai, Zhichao Han, and Xiao-Ming Wu [2018]. “Deeper insights into graph convolutional networks for semi-supervised learning”. In: *Proceedings of The AAAI Conference on Artificial Intelligence*. Vol. 32. 1.
- Li, Shiyang et al. [2019]. “Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting”. In: *Proceedings of the International Conference on Neural Information Processing Systems*. Vol. 32.
- Li, Yaguang et al. [2018]. “Diffusion convolutional recurrent neural network: data-driven traffic forecasting”. In: *Proceedings of the International Conference on Learning Representations*.
- Li, Yujia et al. [2015]. “Gated graph sequence neural networks”. In: *arXiv preprint arXiv:1511.05493*.
- Liao, Shun et al. [2021]. “Deep learning classification of unipolar electrograms in human atrial fibrillation: application in focal source mapping”. In: *Frontiers in Physiology* 12, p. 704122.
- Lipman, Yaron et al. [2022]. “Flow matching for generative modeling”. In: *arXiv preprint arXiv:2210.02747*.
- Lippi, Giuseppe, Fabian Sanchis-Gomar, and Gianfranco Cervellin [2021]. “Global epidemiology of atrial fibrillation: an increasing epidemic and public health challenge”. In: *International Journal of Stroke* 16.2, pp. 217–221.
- Liu, Mingzhe et al. [Feb. 2023]. “PriSTI: A conditional diffusion framework for spatiotemporal imputation”. In: *Proceedings of the IEEE International Conference on Data Engineering (ICDE)*, pp. 1927–1939.

- Liu, Xingchao, Chengyue Gong, and Qiang Liu [2023]. “Flow straight and fast: learning to generate and transfer data with rectified flow”. In: *Proceedings of the International Conference on Learning Representations*.
- Liu, Yukai et al. [2019]. “NAOMI: Non-autoregressive multiresolution sequence imputation”. In: *Proceedings of the International Conference on Neural Information Processing Systems*. Vol. 32.
- Liu, Zijin, Xiang Zhao, and You Song [2025]. “RDPI: A refine diffusion probability generation method for spatiotemporal data imputation”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 39. 12, pp. 12255–12263.
- Lu, Cheng et al. [2025]. “DPM-Solver++: Fast solver for guided sampling of diffusion probabilistic models”. In: *Machine Intelligence Research* 22.4, pp. 730–751.
- Maaten, Laurens Van der and Geoffrey Hinton [2008]. “Visualizing data using t-SNE.” In: *Journal of Machine Learning Research* 9.11.
- Maisel, William H and Lynne Warner Stevenson [2003]. “Atrial fibrillation in heart failure: epidemiology, pathophysiology, and rationale for therapy”. In: *The American Journal of Cardiology* 91.6, pp. 2–8.
- Makridakis, Spyros [1996]. “Forecasting: its role and value for planning and strategy”. In: *International Journal of Forecasting* 12.4, pp. 513–537.
- Mandic, Danilo P [2002]. “Data-reusing recurrent neural adaptive filters”. In: *Neural Computation* 14.11, pp. 2693–2707.
- Mandic, Danilo P and Jonathon Chambers [2001]. *Recurrent neural networks for prediction: learning algorithms, architectures and stability*. John Wiley & Sons, Inc.
- Mandic, Danilo P, Sithan Kanna, and Anthony G Constantinides [2015]. “On the intrinsic relationship between the least mean square and Kalman filters [Lecture Notes]”. In: *IEEE Signal Processing Magazine* 32.6, pp. 117–122.
- Manongi, Ngoda, Joonhyuk Kim, and Seth Goldbarg [2023]. “Dispersion electrogram detection with an artificial intelligence software in redo paroxysmal atrial fibrillation ablation”. In: *HeartRhythm Case Reports* 9.12, pp. 948–953.
- Mao, Huiying, Ryan Martin, and Brian J Reich [2024]. “Valid model-free spatial prediction”. In: *Journal of the American Statistical Association* 119.546, pp. 904–914.
- Marisca, Ivan, Andrea Cini, and Cesare Alippi [2022]. “Learning to reconstruct missing data from spatiotemporal graphs with sparse observations”. In: *Proceedings of the 36th International Conference on Neural Information Processing Systems*, pp. 32069–32082.
- Marques, Antonio G., Santiago Segarra, et al. [2017]. “Stationary Graph Processes and Spectral Estimation”. In: *IEEE Transactions on Signal Processing* 65.22, pp. 5911–5926.
- Martinů, Jan-Matyáš and Petr Šimánek [2024]. “Enhancing weather predictions: super-resolution via deep diffusion models”. In: *Proceedings of the International Conference on Artificial Neural Networks*, pp. 186–197.
- Marwan, Norbert et al. [2007]. “Recurrence plots for the analysis of complex systems”. In: *Physics Reports* 438.5-6, pp. 237–329.

- Mei, Jonathan and José MF Moura [2016]. “Signal processing on graphs: causal modeling of unstructured data”. In: *IEEE Transactions on Signal Processing* 65.8, pp. 2077–2092.
- Meinshausen, Nicolai and Greg Ridgeway [2006]. “Quantile regression forests”. In: *Journal of Machine Learning Research* 7.6.
- Micheli, Alessio [2009]. “Neural network for graphs: a contextual constructive approach”. In: *IEEE Transactions on Neural Networks* 20.3, pp. 498–511.
- Miller, Harvey J and Jiawei Han [May 2009]. *Geographic Data Mining and Knowledge Discovery*. CRC Press. ISBN: 9780429189296.
- Moe, Gordon K, Werner C Rheinboldt, and JA Abildskov [1964]. “A computer model of atrial fibrillation”. In: *American Heart Journal* 67.2, pp. 200–220.
- Mohanty, Sanghamitra et al. [2018]. “Long-term outcome of pulmonary vein isolation with and without focal impulse and rotor modulation mapping: insights from a meta-analysis”. In: *Circulation: Arrhythmia and Electrophysiology* 11.3, e005789.
- Montero-Manso, Pablo and Rob J Hyndman [2021]. “Principles and algorithms for forecasting groups of time series: locality and globality”. In: *International Journal of Forecasting* 37.4, pp. 1632–1653.
- Mou, Liping et al. [2018]. “Lifetime risk of atrial fibrillation by race and socioeconomic status: ARIC study (Atherosclerosis Risk in Communities)”. In: *Circulation: Arrhythmia and Electrophysiology* 11.7, e006350.
- Müller, Luis et al. [2024]. “Attending to graph transformers”. In: *Transactions on Machine Learning Research*.
- Nademanee, Koonlawee et al. [2004]. “A new approach for catheter ablation of atrial fibrillation: mapping of the electrophysiologic substrate”. In: *Journal of the American College of Cardiology* 43.11, pp. 2044–2053.
- Narayan, Sanjiv M and Roy M John [2024]. “Advanced electroanatomic mapping: current and emerging approaches”. In: *Current Treatment Options in Cardiovascular Medicine* 26.4, pp. 69–91.
- Narayan, Sanjiv M, David E Krummen, et al. [2012]. “Treatment of atrial fibrillation by the ablation of localized sources: CONFIRM (Conventional Ablation for Atrial Fibrillation With or Without Focal Impulse and Rotor Modulation) trial”. In: *Journal of the American College of Cardiology* 60.7, pp. 628–636.
- Nattel, Stanley, Brett Burstein, and Dobromir Dobrev [2008]. “Atrial remodeling and atrial fibrillation: mechanisms and implications”. In: *Circulation: Arrhythmia and Electrophysiology* 1.1, pp. 62–73.
- Neifar, Nour et al. [2024]. “DiffECG: A versatile probabilistic diffusion model for ECG signals synthesis”. In: *Proceedings of the IEEE/ACIS International Conference on Software Engineering Research, Management and Applications (SERA)*, pp. 182–188.
- Ng, Fu Siong et al. [2020]. “Toward mechanism-directed electrophenotype-based treatments for atrial fibrillation”. In: *Frontiers in Physiology* 11, p. 987.
- Nguyen, Thao P, Zhilin Qu, and James N Weiss [2014]. “Cardiac fibrosis and arrhythmogenesis: the road to repair is paved with perils”. In: *Journal of Molecular and Cellular Cardiology* 70, pp. 83–91.
- Niculae, Vlad et al. [2023]. “Discrete latent structure in neural networks”. In: *arXiv preprint arXiv:2301.07473*.

- Noble, Denis and Yoram Rudy [2001]. “Models of cardiac ventricular action potentials: iterative interaction between experiment and simulation”. In: *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* 359.1783, pp. 1127–1142.
- Oppenheim, Alan V, Ronald W Schafer, and John R Buck [1999]. *Discrete-Time Signal Processing*. 2nd. Prentice Hall.
- Özgül, Ozan et al. [2023]. “High-density and high coverage composite mapping of repetitive atrial activation patterns”. In: *Computers in Biology and Medicine* 159, p. 106920.
- Pal, Soumyasundar et al. [2021]. “RNN with particle flow for probabilistic spatio-temporal forecasting”. In: *Proceedings of the International Conference on Machine Learning*. Vol. 139. PMLR, pp. 8336–8348.
- Pan, Jiapu and Willis J. Tompkins [1985]. “A real-time QRS detection algorithm”. In: *IEEE Transactions on Biomedical Engineering* 3, pp. 230–236.
- Pandit, Sandeep V and José Jalife [2013]. “Rotors and the dynamics of cardiac fibrillation”. In: *Circulation Research* 112.5, pp. 849–862.
- Papana, Angeliki et al. [2017]. “Financial networks based on Granger causality: a case study”. In: *Physica A: Statistical Mechanics and its Applications* 482, pp. 65–73.
- Paszke, Adam et al. [2019]. “PyTorch: An imperative style, high-performance deep learning library”. In: *Proceedings of the International Conference on Neural Information Processing Systems* 32. Ed. by H. Wallach et al. Curran Associates, Inc., pp. 8024–8035.
- Perraudin, Nathanaël and Pierre Vandergheynst [2017]. “Stationary signal processing on graphs”. In: *IEEE Transactions on Signal Processing* 65.13, pp. 3462–3477.
- Petropoulos, Fotios et al. [2022]. “Forecasting: theory and practice”. In: *International Journal of Forecasting* 38.3, pp. 705–871.
- Petrutiu, Simona, Jason Ng, et al. [2006]. “Atrial fibrillation and waveform characterization”. In: *IEEE Engineering in Medicine and Biology Magazine* 25.6, pp. 24–30.
- Piccini, Jonathan P and Laurent Fauchier [2016]. “Rhythm control in atrial fibrillation”. In: *The Lancet* 388.10046, pp. 829–840.
- Preti, Maria Giulia, Thomas AW Bolton, and Dimitri Van De Ville [2017]. “The dynamic functional connectome: state-of-the-art and perspectives”. In: *NeuroImage* 160, pp. 41–54.
- Pullan, Andrew J, Leo K Cheng, and Martin L Buist [2005]. *Mathematically modelling the electrical activity of the heart: from cell to body surface and back again*. World Scientific.
- Rahman, Shah Atiqur et al. [2015]. “Combining Fourier and lagged k-nearest neighbor imputation for biomedical time series data”. In: *Journal of Biomedical Informatics* 58, pp. 198–207.
- Ramsauer, Hubert et al. [2021]. “Hopfield networks is all you need”. In: *Proceedings of the International Conference on Learning Representations*.
- Rangapuram, Syama Sundar et al. [2018]. “Deep state space models for time series forecasting”. In: *Proceedings of the International Conference on Neural Information Processing Systems* 31, pp. 7785–7794.
- Rantner, LJ et al. [2007]. “Detection of phase singularities in triangular meshes”. In: *Methods of Information in Medicine* 46.06, pp. 646–654.

- Rao, K Venkateswara, A Govardhan, and KV Chalapati Rao [2012]. “Spatiotemporal data mining: issues, tasks and applications”. In: *International Journal of Computer Science and Engineering Survey* 3.1, p. 39.
- Rasul, Kashif et al. [2021]. “Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting”. In: *Proceedings of the International Conference on Machine Learning*, pp. 8857–8868.
- Ravelli, Flavia et al. [2007]. “Deterioration of organization in the first minutes of atrial fibrillation: a beat-to-beat analysis of cycle length and wave similarity”. In: *Journal of Cardiovascular Electrophysiology* 18.1, pp. 60–65.
- Ren, Pengzhen et al. [2021]. “A survey of deep active learning”. In: *ACM Computing Surveys (CSUR)* 54.9, pp. 1–40.
- Rensma, Pieter L et al. [1988]. “Length of excitation wave and susceptibility to reentrant atrial arrhythmias in normal conscious dogs.” In: *Circulation Research* 62.2, pp. 395–410.
- Reuther, Albert et al. [2020]. “Survey of machine learning accelerators”. In: *Proceedings of the IEEE High Performance Extreme Computing Conference (HPEC)*, pp. 1–12.
- Rivera, André et al. [2024]. “Adjunctive low-voltage area ablation for patients with atrial fibrillation: an updated meta-analysis of randomized controlled trials”. In: *Journal of Cardiovascular Electrophysiology* 35.7, pp. 1329–1339.
- Roney, Caroline H et al. [2017]. “Spatial resolution requirements for accurate identification of drivers of atrial fibrillation”. In: *Circulation: Arrhythmia and Electrophysiology* 10.5, e004899.
- Rudy, Yoram [2017]. “Noninvasive ECG imaging (ECGI): mapping the arrhythmic substrate of the human heart”. In: *International Journal of Cardiology* 237, pp. 13–14.
- Rudy, Yoram and John E Burnes [1999]. “Noninvasive electrocardiographic imaging”. In: *Annals of Non-invasive Electrocardiology* 4.3, pp. 340–359.
- Rühling Cachay, Salva et al. [2023]. “DYffusion: A dynamics-informed diffusion model for spatiotemporal forecasting”. In: *Proceedings of the International Conference on Neural Information Processing Systems* 36, pp. 45259–45287.
- Sakata, Kensuke et al. [2024]. “Assessing the arrhythmogenic propensity of fibrotic substrate using digital twins to inform a mechanisms-based atrial fibrillation ablation strategy”. In: *Nature Cardiovascular Research* 3.7, pp. 857–868.
- Salakhutdinov, Ruslan and Andriy Mnih [2008]. “Bayesian probabilistic matrix factorization using Markov chain Monte Carlo”. In: *Proceedings of the 25th International Conference on Machine Learning*, pp. 880–887.
- Salinas, David et al. [2020]. “DeepAR: Probabilistic forecasting with autoregressive recurrent networks”. In: *International Journal of Forecasting* 36.3, pp. 1181–1191.
- Sanchez-Munoz, Juan Jose, Jose Luis Rojo-Alvarez, et al. [2009]. “Spectral analysis of intracardiac electrograms during induced and spontaneous ventricular fibrillation in humans”. In: *Europace* 11.3, pp. 328–331.

- Sandryhaila, Aliaksei and José MF Moura [2013]. “Discrete signal processing on graphs”. In: *IEEE Transactions on Signal Processing* 61.7, pp. 1644–1656.
- Satorras, Victor Garcia, Syama Sundar Rangapuram, and Tim Januschowski [2022]. “Multivariate time series forecasting with latent graph inference”. In: *arXiv preprint arXiv:2203.03423*.
- Sayed, Ali H [2003]. *Fundamentals of adaptive filtering*. John Wiley & Sons.
- Scalzo, Bruno et al. [2023]. “A class of doubly stochastic shift operators for random graph signals and their boundedness”. In: *Neural Networks* 158, pp. 83–88. ISSN: 0893-6080.
- Scarselli, Franco et al. [2008]. “The graph neural network model”. In: *IEEE Transactions on Neural Networks* 20.1, pp. 61–80.
- Scherr, Daniel et al. [2015]. “Five-year outcome of catheter ablation of persistent atrial fibrillation using termination of atrial fibrillation as a procedural endpoint”. In: *Circulation: Arrhythmia and Electrophysiology* 8.1, pp. 18–24.
- Schmidt, Mark, Glenn Fung, and Rómer Rosales [2007]. “Fast optimization methods for ℓ_1 regularization: a comparative study and two new approaches”. In: *Proceedings of the 18th European Conference on Machine Learning (ECML)*, pp. 286–297.
- Seitz, Julien et al. [2022]. “Artificial intelligence software standardizes electrogram-based ablation outcome for persistent atrial fibrillation”. In: *Journal of Cardiovascular Electrophysiology* 33.11, pp. 2250–2260.
- Sen, Rajat, Hsiang-Fu Yu, and Inderjit S Dhillon [2019]. “Think globally, act locally: a deep neural network approach to high-dimensional time series forecasting”. In: *Proceedings of the International Conference on Neural Information Processing Systems* 32.
- Seo, Youngjoo et al. [2018]. “Structured sequence modeling with graph convolutional recurrent networks”. In: *Proceedings of the 25th International Conference on Neural Information Processing*, pp. 362–373.
- Settles, Burr [2009]. *Active learning literature survey*. University of Wisconsin-Madison Department of Computer Sciences.
- Shafipour, Rasoul et al. [2018]. “A directed graph Fourier transform with spread frequency components”. In: *IEEE Transactions on Signal Processing* 67.4, pp. 946–960.
- Shang, Chao and Jie Chen [2021]. “Discrete graph structure learning for forecasting multiple time series”. In: *Proceedings of the International Conference on Learning Representations*.
- Shekhar, Shashi et al. [2015]. “Spatiotemporal data mining: a computational perspective”. In: *ISPRS International Journal of Geo-Information* 4.4, pp. 2306–2338.
- Shi, Lei-Zhi et al. [2015]. “Effect of catheter ablation versus antiarrhythmic drugs on atrial fibrillation: a meta-analysis of randomized controlled trials”. In: *Experimental and Therapeutic Medicine* 10.2, pp. 816–822.
- Shi, Xingjian et al. [2015]. “Convolutional LSTM network: a machine learning approach for precipitation nowcasting”. In: *Proceedings of the International Conference on Neural Information Processing Systems* 28.
- Shuman, David I et al. [2013]. “The emerging field of signal processing on graphs: extending high-dimensional data analysis to networks and other irregular domains”. In: *IEEE Signal Processing Magazine* 30.3, pp. 83–98.

- Smith, Ralph C [2024]. *Uncertainty quantification: theory, implementation, and applications*. SIAM.
- Smyl, Slawek [2020]. “A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting”. In: *International Journal of Forecasting* 36.1, pp. 75–85.
- Song, Y. et al. [2021]. “Score-based generative modeling through stochastic differential equations”. In: *Proceedings of the International Conference on Learning Representations*.
- Song, Yang et al. [2023]. “Consistency models”. In: *Proceedings of the 40th International Conference on Machine Learning*.
- Sperduti, Alessandro [1993]. “Encoding labeled graphs by labeling RAAM”. In: *Proceedings of the International Conference on Neural Information Processing Systems* 6.
- Sperduti, Alessandro and Antonina Starita [1997]. “Supervised neural networks for the classification of structures”. In: *IEEE Transactions on Neural Networks* 8.3, pp. 714–735.
- Stankeviciute, Kamile, Ahmed M Alaa, and Mihaela van der Schaar [2021]. “Conformal time-series forecasting”. In: *Proceedings of the International Conference on Neural Information Processing Systems* 34, pp. 6216–6228.
- Stanković, Ljubiša, Danilo Mandić, et al. [2020a]. “Data analytics on graphs. Part I: Graphs and spectra on graphs”. In: *Foundations and Trends® in Machine Learning* 13.1, pp. 1–157. ISSN: 1935-8237.
- [2020b]. “Data analytics on graphs. Part II: Signals on graphs”. In: *Foundations and Trends® in Machine Learning* 13.2-3, pp. 158–331. ISSN: 1935-8237.
- [2020c]. “Data analytics on graphs. Part III: Machine learning on graphs, from graph topology to applications”. In: *Foundations and Trends® in Machine Learning* 13.4, pp. 332–530. ISSN: 1935-8237.
- Steinhuß, Steven R et al. [2018]. “Effect of a home-based wearable continuous ECG monitoring patch on detection of undiagnosed atrial fibrillation: the mSToPS randomized clinical trial”. In: *JAMA* 320.2, pp. 146–155.
- Stock, James H and Mark W Watson [2002]. “Forecasting using principal components from a large number of predictors”. In: *Journal of the American Statistical Association* 97.460, pp. 1167–1179.
- Stridh, Martin et al. [2001]. “Characterization of atrial fibrillation using the surface ECG: time-dependent spectral properties”. In: *IEEE Transactions on Biomedical Engineering* 48.1, pp. 19–27.
- Sun, Miao et al. [2020]. “Graph-time spectral analysis for atrial fibrillation”. In: *Biomedical Signal Processing and Control* 59, p. 101915.
- Sun, Sophia Huiwen and Rose Yu [2024]. “Copula conformal prediction for multi-step time series prediction”. In: *Proceedings of the International Conference on Learning Representations*.
- Sy, Raymond W., Aravinda Thiagalingam, and Martin K. Stiles [2012]. “Modern electrophysiology mapping techniques”. In: *Heart, Lung and Circulation* 21.6–7, pp. 364–375.
- Taheri, Omid and Sergiy A. Vorobyov [2011]. “Sparse channel estimation with ℓ_p -norm and reweighted ℓ_1 -norm penalized least mean squares”. In: *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2864–2867.
- Tashiro, Yusuke et al. [2021]. “CSDI: Conditional score-based diffusion models for probabilistic time series imputation”. In: *Proceedings of the International Conference on Neural Information Processing Systems*. Vol. 34, pp. 24804–24816.

- Thiagalingam, Aravinda et al. [2004]. “Noncontact mapping of the left ventricle: insights from validation with transmural contact mapping”. In: *Pacing and Clinical Electrophysiology* 27.5, pp. 570–578.
- Trayanova, Natalia A. and Raimond Winslow [2011]. “Whole-heart modeling”. In: *Circulation Research* 108.1, pp. 113–128.
- Trenta, Alessandro, Alessio Gravina, and Davide Bacciu [2025]. “SONAR: Long-range graph propagation through information waves”. In: *Proceedings of the International Conference on Neural Information Processing Systems*.
- Tung, Leslie [1978]. “A bi-domain model for describing ischemic myocardial dc potentials.” PhD thesis. Massachusetts Institute of Technology.
- Umapathy, Karthikeyan et al. [2010]. “Phase mapping of cardiac fibrillation”. In: *Circulation: Arrhythmia and Electrophysiology* 3.1, pp. 105–114.
- Van Rossum, Guido and Fred L. Drake [2009]. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace. ISBN: 1441412697.
- Vandersickel, Nele et al. [2019]. “Directed networks as a novel way to describe and analyze cardiac excitation: directed graph mapping”. In: *Frontiers in Physiology* 10, p. 1138.
- Varidhithai, Thiernithi and Danilo P Mandic [2017]. “On an RLS-like LMS adaptive filter”. In: *Proceedings of the International Conference on Digital Signal Processing*, pp. 1–5.
- Vaswani, Ashish et al. [2017]. “Attention is all you need”. In: *Proceedings of the International Conference on Neural Information Processing Systems*. Vol. 30, pp. 5998–6008.
- Veličković, Petar et al. [2018]. “Graph Attention Networks”. In: *Proceedings of the International Conference on Learning Representations*.
- Verma, Atul et al. [2015]. “Approaches to catheter ablation for persistent atrial fibrillation”. In: *New England Journal of Medicine* 372.19, pp. 1812–1822.
- Vogler, Julia et al. [2015]. “Pulmonary vein isolation versus defragmentation: The CHASE-AF clinical trial”. In: *Journal of the American College of Cardiology* 66.24, pp. 2743–2752.
- Vovk, Vladimir, Alexander Gammerman, and Glenn Shafer [2005]. *Algorithmic learning in a random world*. Vol. 29. Springer.
- Wagner, Patrick et al. [2020]. “PTB-XL, a large publicly available electrocardiography dataset”. In: *Scientific Data* 7.1, pp. 1–15.
- Weiss, James N et al. [2005]. “The dynamics of cardiac fibrillation”. In: *Circulation* 112.8, pp. 1232–1240.
- Wen, Haomin et al. [2023]. “DiffSTG: Probabilistic spatio-temporal graph forecasting with denoising diffusion models”. In: *Proceedings of the 31st ACM International Conference on Advances in Geographic Information Systems*, pp. 1–12.
- Wen, Qingsong et al. [2023]. “Transformers in time series: a survey”. In: *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*.
- Wen, Ruofeng et al. [2017]. “A multi-horizon quantile recurrent forecaster”. In: *arXiv preprint arXiv:1711.11053*.
- Wang, Jay et al. [2022]. “Deblurring via stochastic refinement”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16293–16303.

- Widrow, Bernard and Marcian E. Hoff [1988]. “Adaptive switching circuits”. In: *Neurocomputing: Foundations of Research*. Cambridge, MA, USA: MIT Press, pp. 123–134. ISBN: 0262010976.
- Widrow, Bernard and Samuel D Stearns [1985]. *Adaptive signal processing*. Prentice-Hall.
- Wijffels, Maurits CEF et al. [1995]. “Atrial fibrillation begets atrial fibrillation: a study in awake chronically instrumented goats”. In: *Circulation* 92.7, pp. 1954–1968.
- Winfrey, AT [1994]. “Electrical turbulence in three-dimensional heart muscle”. In: *Science* 266.5187, pp. 1003–1006.
- Winkler, Robert L [1972]. “A decision-theoretic approach to interval estimation”. In: *Journal of the American Statistical Association* 67.337, pp. 187–191.
- Wolf, Philip A et al. [1998]. “Impact of atrial fibrillation on mortality, stroke, and medical costs”. In: *Archives of Internal Medicine* 158.3, pp. 229–234.
- Wong, Kelvin CK et al. [2015]. “No benefit of complex fractionated atrial electrogram ablation in addition to circumferential pulmonary vein ablation and linear ablation: benefit of complex ablation study”. In: *Circulation: Arrhythmia and Electrophysiology* 8.6, pp. 1316–1324.
- Wu, Dongxia et al. [2021]. “Quantifying uncertainty in deep spatiotemporal forecasting”. In: *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pp. 1841–1851.
- Wu, Zonghan et al. [2019]. “Graph wavenet for deep spatial-temporal graph modeling”. In: *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, pp. 1907–1913.
- Xiao, Z., K. Kreis, and A. Vahdat [2022]. “Tackling the generative learning trilemma with denoising diffusion GANs”. In: *Proceedings of the International Conference on Learning Representations*.
- Xie, Sang Michael and Stefano Ermon [2019]. “Reparameterizable subset sampling via continuous relaxations”. In: *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, pp. 3919–3925.
- Xu, Chen, Hanyang Jiang, and Yao Xie [2024]. “Conformal prediction for multi-dimensional time series by ellipsoidal sets”. In: *arXiv preprint arXiv:2403.03850*.
- Xu, Chen and Yao Xie [2023a]. “Conformal prediction for time series”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45.10, pp. 11575–11587.
- [2023b]. “Sequential predictive conformal inference for time series”. In: *Proceedings of the International Conference on Machine Learning*. PMLR, pp. 38707–38727.
- Xu, Maxwell et al. [2022]. “PulseImpute: A novel benchmark task for pulsative physiological signal imputation”. In: *Proceedings of the International Conference on Neural Information Processing Systems*. Vol. 35, pp. 26874–26888.
- Yan, Yi, Ercan E. Kuruoglu, and Mustafa A. Altinkaya [2022]. “Adaptive sign algorithm for graph signal processing”. In: *Signal Processing* 200, p. 108662. ISSN: 0165-1684.
- Yang, Ling et al. [2023]. “Diffusion models: a comprehensive survey of methods and applications”. In: *ACM Computing Surveys* 56.4, pp. 1–39.
- Yin, Xueyan et al. [2022]. “NodeTrans: A graph transfer learning approach for traffic prediction”. In: *arXiv preprint arXiv:2207.01301*.

- Yoon, Jinsung, Daniel Jarrett, and Mihaela Van der Schaar [2019]. “Time-series generative adversarial networks”. In: *Proceedings of the 33rd International Conference on Neural Information Processing Systems* 32.
- Yu, Bing, Haoteng Yin, and Zhanxing Zhu [July 2018]. “Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting”. In: *Proceedings of the International Joint Conference on Artificial Intelligence, IJCAI-18*, pp. 3634–3640.
- Yu, Hsiang-Fu, Nikhil Rao, and Inderjit S Dhillon [2016]. “Temporal regularized matrix factorization for high-dimensional time series prediction”. In: *Proceedings of the International Conference on Neural Information Processing Systems* 29.
- Yun, Donghwan et al. [Nov. 2022]. “Robust R-peak detection in an electrocardiogram with stationary wavelet transformation and separable convolution”. In: *Scientific Reports* 12.1.
- Zaman, Bakht, Luis Miguel Lopez Ramos, et al. [2020]. “Online topology identification from vector autoregressive time series”. In: *IEEE Transactions on Signal Processing* 69, pp. 210–225.
- Zambon, Daniele and Cesare Alippi [2022]. “AZ-whiteness test: a test for signal uncorrelation on spatio-temporal graphs”. In: *Proceedings of the International Conference on Neural Information Processing Systems*.
- [2023]. *Where and how to improve graph-based spatio-temporal predictors*. arXiv: [2302.01701](https://arxiv.org/abs/2302.01701) [cs, stat].
- Zambon, Daniele, Andrea Cini, et al. [2023]. “Graph state-space models”. In: *arXiv preprint arXiv:2301.01741*.
- Zargarbashi, Soroush H, Simone Antonelli, and Aleksandar Bojchevski [2023]. “Conformal prediction sets for graph neural networks”. In: *Proceedings of the International Conference on Machine Learning*. PMLR, pp. 12292–12318.
- Zhang, Tianqi and Ge Guo [2022]. “Graph attention LSTM: a spatiotemporal approach for traffic flow forecasting”. In: *IEEE Intelligent Transportation Systems Magazine* 14.2, pp. 190–196.
- Zhao, Tianyi, Jian Kang, and Lu Cheng [2024]. “Conformalized link prediction on graph neural networks”. In: *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 4490–4499.
- Zheng, Yu et al. [2015]. “Forecasting fine-grained air quality based on big data”. In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 2267–2276.