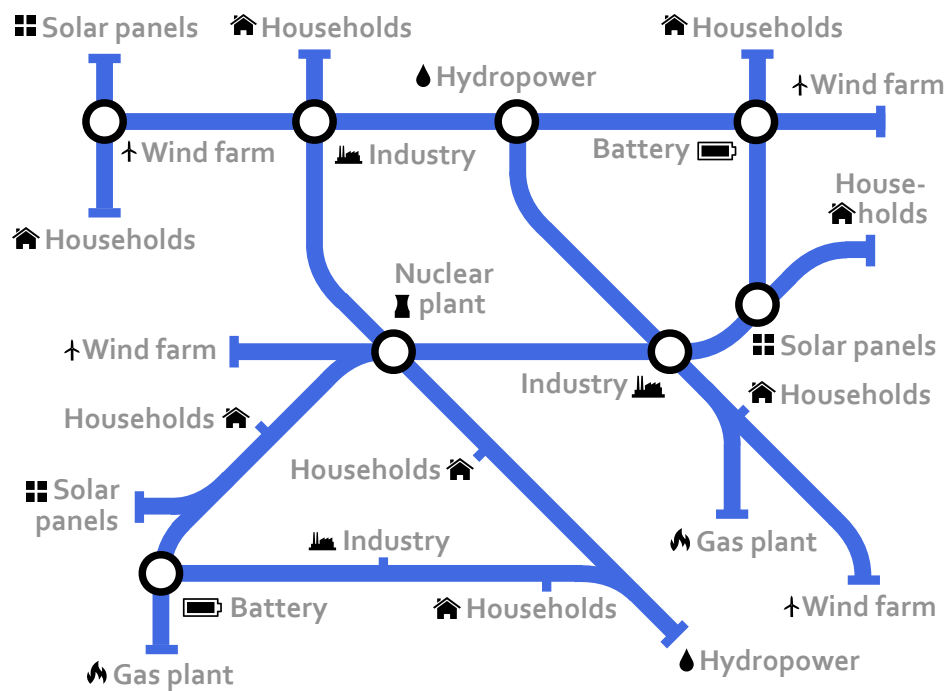




Subsampling for Renewable Electricity System Optimisation

Adriaan Hilbers

Submitted for the degree of Doctor of Philosophy

Department of Mathematics

Imperial College London

2022

Abstract

Decarbonisation efforts are making power (electricity) systems — which require a continuous matching of supply and demand across time and throughout a network known as the grid — increasingly weather-dependent. Historically, this balance was maintained by adjusting dispatchable (mostly fossil fuel) generation. However, as fossil fuels are displaced by weather-dependent renewables such as solar and wind, neither supply nor demand can be fully controlled. This complicates the maintenance of supply-demand balance, known informally as *keeping the lights on*.

Power system planning models are simplified representations of electricity systems used to inform investment strategy, such as whether to build a wind farm, solar farm, gas plant, transmission line or battery. They involve an optimisation problem that minimises the sum of install and subsequent operation costs. Their outputs depend on demand and weather patterns, which enter the model as time series of e.g. demand levels, wind speeds and solar irradiances.

The use of relatively short demand and weather time series (e.g. a single year) in power system planning studies has recently been criticised. This is because many conclusions — such as whether a 100% renewable power system is feasible and affordable — depend on the choice of sample; picking an unrepresentative one provides an unrealistic impression of cost or reliability. At the same time, computing resources usually prevent the consideration of longer samples, since planning problems take too much time or memory to solve.

In this thesis, we introduce subsampling methods to reduce and quantify the impact of demand and weather sampling uncertainty in power system planning models. Our algorithms compress long time series into shorter data sets — enhancing computational efficiency — and provide uncertainty bounds on model outputs. They help prevent incorrect conclusions by (1) allowing the consideration of larger amounts of data, reducing sampling uncertainty, and (2) indicating to users whether their outputs are statistically robust or a result of the particular demand and weather sample.

We combine and extend techniques from the power systems, statistics and optimisation literature. This is because power system planning models involve optimisation problems solved across demand and weather time series, which we view — as is common in statistics — as samples from an underlying distribution. Our contributions are statistical subsampling methods, such as importance sampling and the bootstrap, adapted to power system optimisation problems.

Statement of originality

This thesis summarises a selection of my research at Imperial College London between 2018 and 2022. It is my original work, and any material from other studies is cited and referenced appropriately. The contents of this thesis will not be submitted for any other degree or diploma.

Copyright declaration

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

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe 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.

Acknowledgements

I’ve worked on this research since 2018 and loved it. It would not have been possible without the support of a number of groups and individuals; I’d like to acknowledge these here.

My supervisors, Axel Gandy and David Brayshaw, were fantastic. Axel has a tremendous analytical ability — showing up to meetings with half-completed plots was enough for comments and suggestions that would keep me busy for the next week — and his enthusiasm was infectious. David has a fantastic ability to see the bigger picture, as well as a gift with words; he could effortlessly summarise in one sentence what took me hours to craft into a paragraph. I also appreciate both Axel and David for their focus on deliverables and their selfless encouragement to do what was best for me; this includes their patience during my time out on internships.

This thesis would not exist without the support and community of the Mathematics of Planet Earth Centre for Doctoral Training. I’ll forever hold warm memories of MPE Wednesdays, the CDT space, and Huxley 667. Anna Radomska helped me during my application and is the reason I ended up in this program. Jochen Bröcker and Dan Crisan helped me find my supervisors. Justine Jones spent hours with me on immensely valuable presentation training. Horatio Boedihardjo arranged my internship in Hong Kong, which was a fantastic adventure. These are just a few examples of the help I received.

Imperial College London also provided me with great support. The Research Computing Service has been terrific; their tutorials, help sessions and technical ability felt world-class. Furthermore, among the most enjoyable experiences of the last few years was creating the Coursera course in Tensorflow. I’d like to thank Kevin Webster; the amount of work he put in I cannot imagine, and I loved working with him.

I’d also like to mention the Energy Twitter community. It’s remarkable how much I learned from following the right people. A big thanks to all who participate.

A special thanks to my family, who are a tireless source of support and stability. My parents form a balanced team, encouraging me in both work and my personal life. I’ve also loved the various trips I’ve had with my brother. My family makes my life a lot easier, and for that I am grateful.

I’ve had a great few years in London, both in my work and outside it. Many thanks to everyone who made it such a positive experience.

Contents

1 Introduction	8
1.1 Power system planning models	8
1.2 Power systems, optimisation and statistics	10
1.3 The consequences of computational cost	10
1.4 Contributions	10
1.5 Thesis outline	12
2 Scientific background	13
2.1 Power systems	13
2.2 Power system models	14
2.3 Power system planning models	14
2.3.1 Operation problem	15
2.3.2 Planning problem	16
2.3.3 Interpretation as functions	17
2.4 Computational complexity	18
2.5 Demand and weather uncertainty	18
2.6 Stochastic programming	20
2.6.1 Sample average approximation	20
2.6.2 Two-stage problems	21
2.7 Planning by stochastic programming	21
2.7.1 Mathematics	22
2.7.2 Remarks	23
3 Power system models for time series analysis	24
3.1 Overview	24
3.2 Open-source software package	26
3.3 Solvers and computational cost	27
3.4 Decomposition methods	28
3.4.1 Operation problem	28
3.4.2 Planning problem	30
3.5 Mathematical details	33
3.5.1 Technologies	33
3.5.2 Time series	34
3.5.3 Mathematical formulation	35

4	Efficient design estimates without storage	38
4.1	Background	38
4.1.1	Context: time series aggregation	38
4.1.2	Mathematical background	39
4.1.3	Literature review	42
4.1.4	Conclusions and implications	43
4.2	Contribution	45
4.3	Methods	45
4.3.1	Overview and intuition	45
4.3.2	Algorithm: <i>importance subsampling</i>	47
4.3.3	Remarks	48
4.4	Simulation studies	49
4.4.1	Overview	49
4.4.2	Setup	50
4.4.3	Results	51
4.4.4	Discussion	53
4.5	Conclusions	54
5	Efficient design estimates with storage	56
5.1	Background	56
5.1.1	Context: time series aggregation	56
5.1.2	Mathematical background	58
5.1.3	Literature review	61
5.1.4	Conclusions and implications	62
5.2	Contribution	63
5.3	Methods	63
5.3.1	Overview and intuition	63
5.3.2	Algorithm: <i>storage importance subsampling</i>	64
5.3.3	Remarks	65
5.4	Simulation studies	66
5.4.1	Overview	66
5.4.2	Setup	67
5.4.3	Results	69
5.4.4	Discussion	71
5.5	Conclusions	72
6	Efficient uncertainty quantification	74
6.1	Background	74
6.1.1	Context: uncertainty quantification	74

6.1.2 Literature review	75
6.1.3 Conclusions and implications	77
6.2 Contribution	78
6.3 Methods	78
6.3.1 Overview	78
6.3.2 Algorithm: m -out-of- n bootstrap for $\sigma_{\hat{D}}$	79
6.3.3 Remarks	79
6.3.4 Justification and a test for validity	80
6.4 Simulation studies	81
6.4.1 Setup	81
6.4.2 Quantifying output uncertainty	83
6.4.3 Determining required sample length	85
6.4.4 Verifying the method's consistency	86
6.4.5 Discussion	88
6.5 Conclusions	89
7 Conclusions	90
7.1 Context	90
7.2 Contributions	90
7.3 Implications	91
7.4 Extensions	92
A Additional scientific background	93
A.1 Monte Carlo methods	93
A.1.1 Expected values	94
A.1.2 Other use cases	95
A.2 Importance sampling	96
A.3 Bootstrap methods	97
A.3.1 Machinery	98
A.3.2 The m -out-of- n bootstrap	99
A.4 Sample average approximation: theory	99
A.4.1 M-estimators	100
A.4.2 Sample average approximations	100
Nomenclature	102
Bibliography	103

1. Introduction

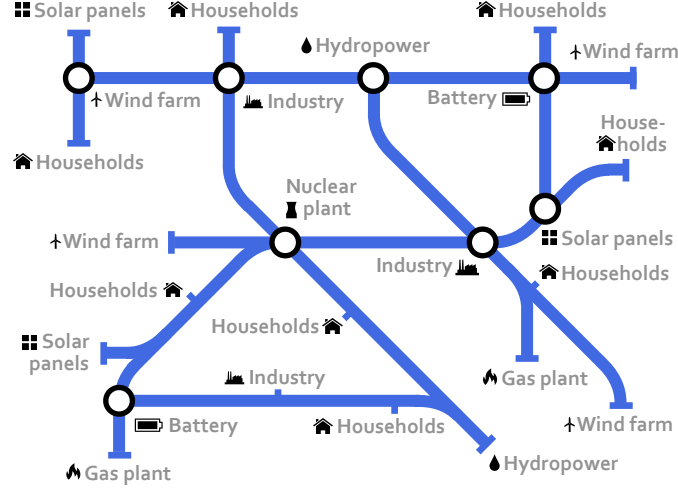


Figure 1.1: Power system planning models use simplified representations of electrical grids, including generators, consumers and transmission lines.

In this section, we briefly introduce this thesis’ research context and our contributions. We discuss power system planning models (Section 1.1), their link to optimisation and statistics (Section 1.2) and their computational cost (Section 1.3). We also summarise our contributions (Section 1.4) and this document’s structure (Section 1.5). For detailed scientific background, including a literature review, see Chapter 2.

1.1 Power system planning models

A *power system* is the infrastructure built around electricity: generation, transmission and storage technologies across a network known as a *grid*. Power system operation requires a continuous matching of supply and demand, both in time and space; maintaining it is referred to informally as *balancing the grid* [Stoft, 2002].

Decarbonisation efforts in the face of climate change affect both sides of the supply-demand balance. On the demand side, new sectors, such as transport and heating, are increasingly electrified. On the supply side, fossil fuels such as coal and gas are replaced by carbon-free generation, notably variable renewables such as solar and wind whose generation depends on the weather. In this case, neither supply nor demand can be fully controlled and *security of supply* — the ability to balance

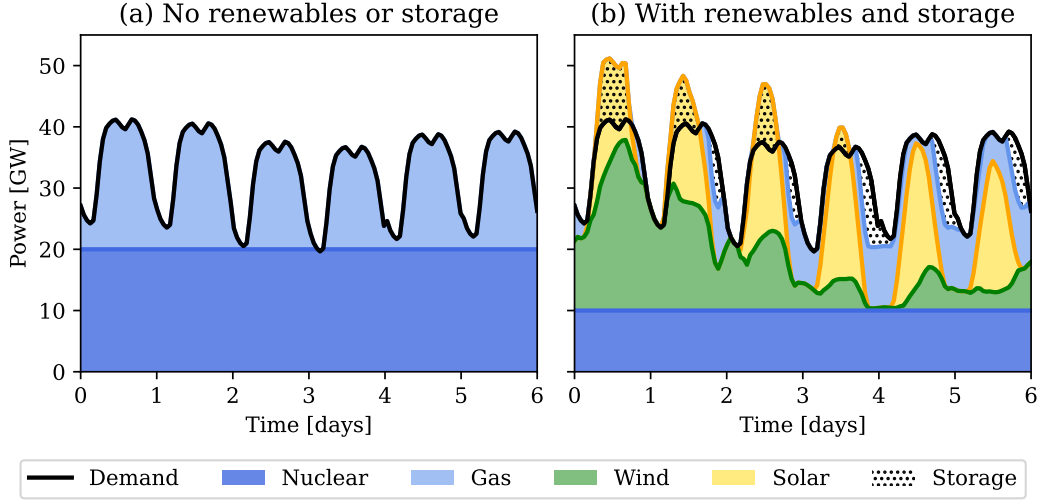


Figure 1.2: Supply and demand balance in single-region power system. Demand is met continuously by a combination of generation technologies. Storage charges using excess generation and discharges when demand exceeds supply.

the grid across a range of scenarios — is more difficult to maintain [Jenkins et al., 2021](#).

Power system planning models are simplified representations of electricity systems used to inform their design and operation [Pfenninger et al., 2014](#). In this thesis, we consider a subclass: *capacity expansion planning* models, which find the system design that minimises the total (sum of install and subsequent operation) cost across a simplified grid (Figure [1.1](#)) subject to constraints, such as demand being met continuously as in Figure [1.2](#) [Merrick, 2016](#).

The planning models we consider are solved across demand and weather time series such as hourly demand levels, wind speeds or solar irradiances. In this thesis, we view models as functions of this data:

$$\hat{\mathbf{D}} = \text{PSM}_{\text{plan}}((\xi_t)_{t \in \mathcal{T}}) \quad (1.1)$$

where ξ_t is the demand and weather data at time t , $\mathcal{T}$ indexes the time steps and $\hat{\mathbf{D}}$ is the system design that minimises total cost: a vector of generation, transmission and storage capacities across the grid^{[1](#)}.

¹In this thesis, we consider *build from scratch* models in which the investment is the full system, but ideas generalise naturally to investments into existing systems.

1.2 Power systems, optimisation and statistics

The equation (1.1) provides a conceptual framework. The system design $\hat{\mathbf{D}} = \hat{\mathbf{D}}((\boldsymbol{\xi}_t)_{t \in \mathcal{T}})$ minimises the total cost across the demand and weather time series $(\boldsymbol{\xi}_t)_{t \in \mathcal{T}}$, which we view as a sample from some (unknown) demand and weather distribution. We then view $\hat{\mathbf{D}}$ as an estimate for $\mathbf{D}$, the (unknown) optimal design that minimises *expected* costs under this distribution.

We consider $\hat{\mathbf{D}}$ from three perspectives: power systems, optimisation and statistics. From a power systems perspective, $\hat{\mathbf{D}}$ is an estimate of the system design $\mathbf{D}$ that minimises expected total costs. From an optimisation perspective, $\hat{\mathbf{D}}$ is the minimiser of a cost function given constraints based on time series data (see Section 2.3). From a statistics perspective, $\hat{\mathbf{D}}$ is a statistical *estimator* for $\mathbf{D}$ as is common in *Monte Carlo* methods (Appendix A.1). These perspectives lead to an interpretation of planning models as *stochastic programming* as we discuss in Section 2.7.

1.3 The consequences of computational cost

Computing resources typically restrict planning problems to be solved across relatively short samples (e.g. a few weeks or months). This is because (1) realistic modelling of power systems requires a detailed representation of technologies and the grid [Poncelet et al., 2016] and (2) solution times scale quickly in the number of decision variables, constraints and time steps [Cao et al., 2019].

The use of short samples creates a risk of incorrect conclusions. For example, the feasibility and affordability of 100% renewable systems (without nuclear, biomass or carbon capture and storage) depends strongly on the choice of demand and weather data [Shaner et al., 2018; Zappa et al., 2019]. However, the most obvious solution to reducing this sampling uncertainty — solving the planning model across longer time series — is usually computationally unfeasible. See Sections 2.4 and 2.5 for a more detailed discussion on computational cost and sampling uncertainty.

1.4 Contributions

This thesis considers demand and weather sampling uncertainty in power system planning models. Figure 1.3(a) illustrates the problem. We estimate optimal design $\mathbf{D}$ (which minimises expected costs) by $\hat{\mathbf{D}}$, obtained from a planning model as in (1.1). Estimates $\hat{\mathbf{D}}$ may have a large variability and bias (compared to $\mathbf{D}$) across

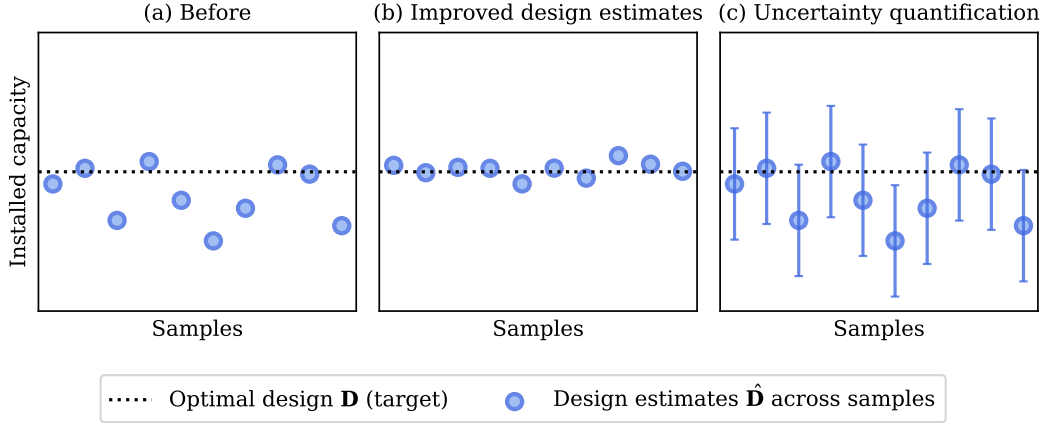


Figure 1.3: Contributions. (a) Design estimates $\hat{\mathbf{D}}$ have high bias and variance compared to true optimum $\mathbf{D}$ (which minimises the expected costs under uncertainty). Our methods (b) improve point estimates and (c) quantify uncertainty. Each sample is plotted at different x -value; y -values are design variables, e.g. installed wind generation capacity.

samples $(\xi_t)_{t \in \mathcal{T}}$ of demand and weather data, meaning decisions motivated by such simulations may be incorrect in the long term.

In this thesis, we introduce algorithms to mitigate the impact of this uncertainty by subsampling demand and weather time series. These subsamples are shorter than the originals — reducing computational cost — and carefully chosen to induce as little approximation error as possible. Our contributions combine and expand techniques from power systems, optimisation and statistics literature, and can be classified into two groups as we describe below.

One class of contributions improves point estimates as shown in Figure 1.3(b). These estimates have lower variability and bias, providing more accurate estimates of $\mathbf{D}$. They are determined using samples that are designed to include relevant extreme events, which are disproportionately important to accurate design estimates. We introduce this method in Chapter 4 and extend it to models with storage technologies in Chapter 5.

Another contribution is an uncertainty quantification scheme as illustrated in Figure 1.3(c). It allows the construction of confidence intervals for design estimates and indicates a range of possible values for $\hat{\mathbf{D}}$ or $\mathbf{D}$. The method conducts Monte Carlo uncertainty analysis on shorter demand and weather time series, obtained from the original via stratified block bootstrapping. We describe this method in Chapter 6.

Contributed papers

Our contributions have been published in the following academic papers:

1. Chapter 4: Hilbers et al., 2019. AP Hilbers, DJ Brayshaw, A Gandy. Importance subsampling: improving power system planning under climate-based uncertainty. *Applied Energy*, 2019.
2. Chapter 4: Hilbers et al., 2020. AP Hilbers, DJ Brayshaw, A Gandy. Importance subsampling for power system planning under multi-year demand and weather uncertainty. In *2020 IEEE International Conference on Probabilistic Methods Applied to Power Systems*, 2020.
3. Chapter 5: Hilbers et al., 2022. AP Hilbers, DJ Brayshaw, A Gandy. Reducing climate risk in energy system planning: a posteriori time series aggregation for models with storage. *Applied Energy*, 2023.
4. Chapter 6: Hilbers et al., 2021. AP Hilbers, DJ Brayshaw, A Gandy. Efficient quantification of the impact of demand and weather uncertainty in power system models. *IEEE Transactions on Power Systems*, 2021.

Contributed software

The power system planning problems used in our experiments form part of a software package that we make publicly available. A high-level API, released as a `python` package, helps study the impact of demand and weather (time series) on power systems. We also release a tutorial that has been used as an introduction to energy system modelling. A repository containing the package, model files, time series data and documentation is available at doi:[10.5281/zenodo.7025098](https://doi.org/10.5281/zenodo.7025098). For a detailed discussion, see Chapter 3.

1.5 Thesis outline

This document is structured as follows. Chapter 2 provides scientific background and places this thesis within the literature. Chapter 3 discusses our power system models and software. Chapter 4 introduces the method for improved point estimates and Chapter 5 extends it to models with storage technologies. Chapter 6 introduces the method for uncertainty quantification. Chapter 7 provides concluding remarks. Appendix A discusses mathematical and statistical theory.

2. Scientific background

In this chapter, we discuss the scientific background this thesis builds on. We introduce power systems (Section 2.1), power system models (Section 2.2) and the planning models we study (Section 2.3). We discuss computational complexity (Section 2.4), demand and weather uncertainty (Section 2.5) and stochastic programming (Section 2.6). Finally, we explain how power system planning problems may be viewed as sample average approximations to stochastic programming problems. We discuss only material essential to the understanding of this document; Appendix A provides additional background and details.

2.1 Power systems

A *power system* is the infrastructure around electricity, including generators, storage and the transmission/distribution grid. Its operation is constrained by electrical physics and requires a continuous matching of supply and demand, both temporally (each point in time) and spatially (each grid location). Practically, this means *balancing the grid* via corrective actions such as adjusting generation levels, transmitting electricity or, in extreme cases, curtailing demand [Stoft, 2002, Xu, 2016].

Efforts to mitigate climate change are transforming power systems. The 2015 Paris Climate Agreement and subsequent decarbonisation plans require a rapid decrease in greenhouse gas emissions [Climate Action Tracker, 2020]. For example, the United Kingdom is committed to a reduction of annual carbon emissions (compared to 1990 levels) of 78% by 2035 and *net zero* by 2050 [Committee on Climate Change, 2020]. A common roadmap involves decarbonising electricity generation and electrifying other sources of emissions, such as transport and heating [Tsiropoulos et al., Jenkins et al., 2021]. In most decarbonisation scenarios, the proportion of energy coming from variable renewable sources such as solar and wind increases considerably [National Grid, 2021]. This creates a paradigm in which neither electricity demand nor supply can be fully controlled, complicating the maintenance of grid balance under a range of demand and weather scenarios.

2.2 Power system models

Power system models (PSMs) simulate how electricity is generated, transmitted and consumed across a simplified representation of a grid (Figure 1.1). They allow scenarios such as a fully renewable system or electrified transport and heating to be examined. PSMs are a subclass of *energy system models*, which include e.g. oil, gas, and agriculture as well as electricity.

PSMs can be classified on a number of criteria. Two key factors are spatiotemporal resolution and techno-economic details; some models have high resolution and intricate technical detail (considering e.g. the physics of alternating current electricity) while others have more simplified representations. Another key differentiator is mathematical formulation, which include *simulation* (follow a set of rules to determine system behaviour), *optimisation* (minimise a cost function under constraints), *agent-based* (simulate independent participants, such as private firms) and many more.

A number of review papers discuss energy and power system modelling. Pfenniger et al. [2014] and Lopion et al. [2018] provide a high-level overview of model types and research challenges. Lund et al. [2017] discuss trends in simulation- and optimisation-based energy system modelling. Bazmi and Zahedi [2011] discuss the role of optimisation methods in PSMs.

2.3 Power system planning models

In this thesis, we consider *power system planning* models, also known as *capacity expansion planning* models [Merrick, 2016]. These are optimisation-based PSMs widely used in decision support for infrastructure investments [Loulou et al., 2004, Strachan and Li, 2021]. Mathematically, they determine cost-optimal system design (generation, transmission and storage capacities) by minimising the sum of install (build) cost and subsequent operation cost while meeting demand.

Each planning model has an associated *operation* model, in which we fix the system design and minimise only the operation costs; this is sometimes called the *production cost model* [Teichgraeber and Brandt, 2022]. In the next two sections, we introduce the operation problem (the optimisation problem associated with the *operation* model), followed by the planning problem that we study throughout this thesis.

2.3.1 Operation problem

The *operation* problem minimises the operation cost: the cost to meet demand across the grid by generating and transmitting electricity. In region r at time t , let $\text{gen}_{i,r,t}$ be the generation from technology i , $\text{tr}_{r,r',t}$ be the transmission to region r' , and $\text{ch}_{r,t}$ be the storage charging rate, all measured in MW. Let $\mathcal{I}$, $\mathcal{R}$ and $\mathcal{T}$ be the set of generation technologies, grid regions and time steps respectively. Then $\mathbf{O}_t := [\text{gen}_{i,r,t}, \text{tr}_{r,r',t}, \text{ch}_{r,t} \mid i \in \mathcal{I}; r, r' \in \mathcal{R}]$ describes the system *operation* at time t , and the operation problem is

$$\min_{(\mathbf{O}_t)_{t \in \mathcal{T}}} \underbrace{\sum_{t \in \mathcal{T}} \text{cost}_{\text{operate}}(\mathbf{O}_t)}_{\text{total operation cost}} \quad (2.1)$$

subject to

$$\sum_{i \in \mathcal{I}} \text{gen}_{i,r,t} + \sum_{r' \in \mathcal{R}} \text{tr}_{r',r,t} = d_{r,t} + \text{ch}_{r,t} \quad \text{for all } r, t \quad (2.2)$$

$$\text{tr}_{r,r',t} + \text{tr}_{r',r,t} = 0 \quad \text{for all } r, r', t \quad (2.3)$$

$$\text{sto}_{r,0} = 0 \quad \text{for all } r \quad (2.4)$$

$$\text{sto}_{r,t+1} = (1 - l^{\text{sto}}) \text{sto}_{r,t} + \begin{cases} e^{\text{sto}} \text{ch}_{r,t} & \text{if } \text{ch}_{r,t} \geq 0 \\ \frac{1}{e^{\text{sto}}} \text{ch}_{r,t} & \text{if } \text{ch}_{r,t} < 0 \end{cases} \quad \text{for all } r, t \quad (2.5)$$

$$0 \leq \text{gen}_{i,r,t} \leq \text{cap}_{i,r}^{\text{gen}} \lambda_{i,r,t} \quad \text{for all } i, r, t \quad (2.6)$$

$$|\text{tr}_{r,r',t}| \leq \text{cap}_{r,r'}^{\text{tr}} + \text{cap}_{r',r}^{\text{tr}} \quad \text{for all } r, r', t \quad (2.7)$$

$$0 \leq \text{sto}_{r,t} \leq \text{cap}_r^{\text{sto}} \quad \text{for all } r, t \quad (2.8)$$

$$(\text{optional additional constraints}) \quad (2.9)$$

The function $\text{cost}_{\text{operate}}$ is typically the generation cost. In region r , the variables $\text{cap}_{i,r}^{\text{gen}}$, $\text{cap}_{r,r'}^{\text{tr}}$ and $\text{cap}_r^{\text{sto}}$ are the installed capacity of generation technology i (MW), transmission to region r' (MW) and energy storage (MWh) respectively. (2.2) indicates that the sum of generation and transmission into a region equals demand plus storage charging. (2.3) is the transmission balance. (2.4) specifies empty storage at the start of a simulation and (2.5) indicates how storage levels change with storage (dis)charging efficiency e^{sto} and self-loss l^{sto} . (2.6) and (2.7) constrain generation and transmission levels to their capacities. Generation is capped by installed capacity times *generation potential* $\lambda_{i,r,t} \in [0, 1]$. For *dispatchable* (e.g. fossil fuel or

nuclear) generation, $\lambda_{i,r,t} \equiv 1$; a 100MW plant generates at most 100MW (we ignore outages). For variable renewables, $\lambda_{i,r,t}$ depends on the weather. For example, for wind, $\lambda_{i,r,t} = 0.1$ and $\lambda_{i,r,t} = 0.9$ correspond to times of low and high wind speeds respectively. (2.8) constrains storage levels to storage energy bounds. We constrain storage *energy* (MWh) but not instantaneous *power* (dis)charge rates (MW). Optional constraints (2.9) may include *ramping* constraints on rates of change of generation levels or minimum generation levels when plants are turned on.

Each problem is solved given a sample of demand and weather data $(\xi_t)_{t \in \mathcal{T}}$, with $\xi_t = [d_{r,t}, \lambda_{i,r,t} \mid i \in \mathcal{I}; r \in \mathcal{R}]$. Demand levels $d_{r,t}$ appear in (2.2) and weather determines the generation potentials $\lambda_{i,r,t}$ in (2.6).

For this problem, generation, transmission and storage capacities ($\text{cap}_{i,r}^{\text{gen}}, \text{cap}_{r,r'}^{\text{tr}}$ and $\text{cap}_r^{\text{sto}}$) are fixed, corresponding to a fixed system design. Such problems are sometimes called *economic dispatch* problems, or *unit commitment economic dispatch* if power plants have a binary on/off state.

2.3.2 Planning problem

The *planning* problem determines the design that minimises the sum of install (build) and subsequent operation cost. Let $\mathbf{D} := [\text{cap}_{i,r}^{\text{gen}}, \text{cap}_{r,r'}^{\text{tr}}, \text{cap}_r^{\text{sto}} \mid i \in \mathcal{I}; r, r' \in \mathcal{R}]$ be the power system design (installed capacities of generation, transmission and storage technologies). The planning problem is

$$\min_{\mathbf{D} \geq \mathbf{0}, (\mathbf{O}_t)_{t \in \mathcal{T}}} \left[\underbrace{|\mathcal{T}| \times \text{cost}_{\text{install}}(\mathbf{D})}_{\text{normalised install cost}} + \underbrace{\sum_{t \in \mathcal{T}} \text{cost}_{\text{operate}}(\mathbf{O}_t)}_{\text{total operation cost}} \right] \quad (2.10)$$

subject to

$$(\text{operational constraints (2.2)-(2.9)}) \quad (2.11)$$

$$(\text{optional design constraints}). \quad (2.12)$$

Usually, the only constraint on $\mathbf{D}$ is that all capacities (its components) are non-negative. Examples of optional design constraints include building capacity in units of a certain size (corresponding to whole numbers of generation plants) or restrictions on transmission capacities.

Let $\hat{\mathbf{D}}$ and $(\hat{\mathbf{O}}_t)_{t \in \mathcal{T}}$ be the minimisers of this problem when they are unique. The planning problem is the operation problem except that the system design $\mathbf{D}$ is deter-

mined instead of fixed, and its cost is added to the objective function. We optimise over both $\mathbf{D}$ and the subsequent operation $(\mathbf{O}_t)_{t \in \mathcal{T}}$ but are interested primarily in $\mathbf{D}$. To normalise the install and operation costs to the same temporal scale, we let $\text{cost}_{\text{install}}(\mathbf{D})$ be the install cost per time step of plant lifetime and multiply by $|\mathcal{T}|$ (the number of time steps).

The optimisation is a trade-off between install and operation costs. The design $\mathbf{D}$ constrains generation transmission and storage levels in (2.6)-(2.8). Higher capacities, with a higher install cost, loosen the bounds on operational constraints, meaning a lower operation cost. Conversely, lower capacity levels reduce install cost but increase operational cost.

In this thesis, we consider *build from scratch* power system planning models in which the entire system design is optimised. Similar models determine investments into an *existing* system; these have an identical formulation with $\mathbf{D} = \mathbf{D}_0 + \mathbf{D}_I$, where $\mathbf{D}_0$ is fixed existing capacity and $\mathbf{D}_I$ is new infrastructure.

2.3.3 Interpretation as functions

It is useful to interpret the planning problem as a function from a demand and weather time series $(\xi_t)_{t \in \mathcal{T}}$ to the associated optimal system design $\hat{\mathbf{D}}$:

$$\hat{\mathbf{D}} = \text{PSM}_{\text{plan}}((\xi_t)_{t \in \mathcal{T}}) \quad (2.13)$$

where $\hat{\mathbf{D}}$ is determined by solving the planning problem (2.10)-(2.12). We may interpret the operation problem similarly:

$$(\hat{\mathbf{O}}_t)_{t \in \mathcal{T}} = \text{PSM}_{\text{operate}}((\xi_t)_{t \in \mathcal{T}} \mid \mathbf{D}) \quad (2.14)$$

where the argument $\mathbf{D}$ is the fixed design for which the operation is determined.

The planning problem is solved deterministically given a sample $(\xi_t)_{t \in \mathcal{T}}$ of demand and weather data. Different samples (e.g. years) are possible, leading to different designs. In Section 2.7, we explain how such design estimates $\hat{\mathbf{D}}$ are *sample average approximations* to the design $\mathbf{D}$ that minimises *expected* costs under *all* possible demand and weather samples (Section 2.5).

In this thesis, we concentrate on uncertainty on optimal design estimates $\hat{\mathbf{D}}$ coming from planning models, considering how much they vary under different samples $(\xi_t)_{t \in \mathcal{T}}$ as in equation (2.13). For example, we might consider how much estimates of cost-optimal wind generation capacity vary across different weather years. We

do not study operational uncertainties. For example, even for a fixed design, operational variables such as generation levels and costs vary year-to-year as determined via the operational problem (2.14). We introduce operational problems since they appear in some of our methods, but we primarily study the uncertainty on design estimates from planning problems.

2.4 Computational complexity

The computing time and memory required to solve planning models can be significant [Lopion et al., 2018, Ringkjøb et al., 2018, Kotzur et al., 2021]. For example, for our *MILP* model, a yearly simulation with hourly time steps takes between one and four days (see Section 3.3 for details).

One driver of computational cost is the number of decision variables and constraints. Planning models often contain many regions and technologies, and capturing demand and weather variability requires a high temporal resolution. [Poncelet et al., 2016, Merrick 2016 and Collins et al., 2017] find that models with time steps longer than one hour underestimate costs and overestimate the reliability of renewables. The large number of technologies, regions and time steps increase model complexity [DeCarolis et al., 2017, Kotzur et al., 2021].

Another driver is that algorithms used to solve planning models scale quickly in the number of decision variables and constraints. [Goderbauer et al., 2019] show that planning problems with integer variables (such as the *MILP* model in Section 3.1) are NP-hard. However, even for models with only continuous variables, solution times increase much faster than linearly (Section 3.3); one explanation is that the system design $\mathbf{D}$ contains *complicating* decision variables [Cao et al., 2019] that appear in each time step and prevent temporal decomposition into subproblems.

Methods to reduce computational cost in power system planning models form an active field of research, as reviewed by [Cao et al., 2019]. A particularly common approach is *time series aggregation* — compressing long time series into shorter samples — which we review in Sections 4.1.1 and 5.1.1.

2.5 Demand and weather uncertainty

The impact of weather in power system models is increasing [Bloomfield et al., 2021]. On the supply side, weather-dependent renewables represent an increasing proportion of generation. On the demand side, electrification of heating and cooling

is increasing the variability and weather-dependence of demand patterns [National Grid, 2021].

Optimal design estimates from power system planning models are hence subject to *demand and weather uncertainty*. In (2.13), $\hat{\mathbf{D}}$ is calculated deterministically across the time series $(\xi_t)_{t \in \mathcal{T}}$, which we view as a sample from some underlying demand and weather distribution. For example, each year of data produces an estimate of cost-optimal system design $\hat{\mathbf{D}}$. We use the term *demand and weather uncertainty* to refer to this sampling uncertainty.

A number of investigations consider the impact of demand and weather uncertainty on planning and/or operation models. They typically run the same model across different time periods (usually years) and consider e.g. cost-optimal system design, costs and emissions. Investigations include the United Kingdom [Bloomfield et al., 2016, Pfenninger, 2017, Staffell and Pfenninger, 2018, Zeyringer et al., 2018], Germany [Wohland et al., 2018], Portugal [Amorim et al., 2020], Europe [Collins et al., 2018, Ravestein et al., 2018], Texas [Bothwell and Hobbs, 2018, Kumler et al., 2019], Hawaii [Bryce et al., 2018] and the United States [Shaner et al., 2018]. They find large variability in model outputs across demand and weather samples, especially for systems with significant levels of renewable generation ($>10\%$). For example, both [Bloomfield et al., 2016] and [Zeyringer et al., 2018] find that optimal generation capacity levels for specific technologies in the United Kingdom range anywhere from 0% to 250% of their mean value across different years. [Bothwell and Hobbs, 2018] note that systems optimised for individual years miss environmental targets or have high rates of unmet demand in others.

This variability creates a risk for decision-makers. While users are unlikely to trust precise numerical values derived from planning models (since these are subject to various other uncertainties, e.g. financing or future electricity costs), demand and weather uncertainty can also impact qualitative decisions. For example, whether a highly ($>80\%$) renewable power system is feasible and affordable depends heavily on the choice of demand and weather data [Shaner et al., 2018, Zappa et al., 2019]. In this way, the naïve consideration of short demand and weather samples may lead to incorrect interpretations of e.g. the reliability of renewables, whether fully renewable power is possible, and whether battery storage alone is sufficient as backup power.

Uncertainty of this type appears in any capital planning decision based on a sample of data. Such exercises do not give the quantity actually desired (since future data will differ from that optimised across), but can be a useful proxy if future events occur according to the same distribution. Samples can be from historic data (with

a stationarity assumption justifying their use for forward-looking decisions) or from e.g. a simulation model. We study the sampling uncertainty coming from the use of a finite dataset. For more details on such stochastic programming problems, see [Birge and Louveaux, 2011, Chapter 1].

2.6 Stochastic programming

Power system planning models have a conceptual link to *stochastic programming*, which we introduce briefly in this section. We discuss the link to planning models in Section 2.7.

Stochastic programming minimises an objective function under uncertainty from a random variable [Shapiro et al., 2009]. Consider a real-valued function $F(\mathbf{x}, \boldsymbol{\xi})$ that depends on a decision variable $\mathbf{x}$ and a random variable $\boldsymbol{\xi}$ with probability measure $\mathbb{P}$, and let $\mathbb{E}(F(\mathbf{x}, \boldsymbol{\xi})) = \int F(\mathbf{x}, \boldsymbol{\xi}) d\mathbb{P}(\boldsymbol{\xi})$ be its expected value. We consider problems of the form

$$\min_{\mathbf{x} \in \mathcal{X}} \mathbb{E}(F(\mathbf{x}, \boldsymbol{\xi})) \quad (2.15)$$

with $\mathcal{X}$ independent of $\boldsymbol{\xi}$. Both $\mathbf{x}$ and $\boldsymbol{\xi}$ are vector-valued in general. When applicable, we denote this problem's minimum by θ^* and minimiser by $\mathbf{x}^*$.

2.6.1 Sample average approximation

The *sample average approximation* to (2.15) minimises the objective function across a *sample* $\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n$ drawn from $\mathbb{P}$:

$$\min_{\mathbf{x} \in \mathcal{X}} \left[\frac{1}{n} \sum_{i=1}^n F(\mathbf{x}, \boldsymbol{\xi}_i) \right]. \quad (2.16)$$

Let $\hat{\theta}_n = \hat{\theta}_n(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n)$ and $\hat{\mathbf{x}}_n = \hat{\mathbf{x}}_n(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n)$ be its minimum and minimiser respectively, when they are unique.

The sample average approximation (2.16) is useful since, in many settings, its solution converges to the same value as that of the underlying stochastic programming problem (2.15) [Homem-de Mello and Bayraksan, 2014]. By the central limit theorem, $\frac{1}{n} \sum_{i=1}^n F(\mathbf{x}, \boldsymbol{\xi}_i) \xrightarrow{\mathbb{P}} \mathbb{E}(F(\mathbf{x}, \boldsymbol{\xi}))$ pointwise for any $\mathbf{x}$ for which $\text{Var}(F(\mathbf{x}, \boldsymbol{\xi}))$ is finite (see Appendix A.1). Furthermore, in many settings, the solutions to (2.16) are *consistent*, with $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta^*$ and $\hat{\mathbf{x}}_n \xrightarrow{\mathbb{P}} \mathbf{x}^*$ as $n \rightarrow \infty$. A discussion on consistency of sample average approximations is provided in Appendix A.4; the key conditions are a uniform convergence of the objective function and the optimal value $\mathbf{x}^*$ lying

away from the boundary of its feasible region.

2.6.2 Two-stage problems

Two-stage problems [Dantzig and Glynn, 1990] decompose into the sum of a first-stage cost independent of ξ and a second-stage subproblem depending on both $\mathbf{x}$ and the random variable ξ . In this case,

$$F(\mathbf{x}, \xi) = F_{\mathbf{x}}(\mathbf{x}) + F_{\mathbf{y}}^*(\mathbf{x}, \xi) \quad (2.17)$$

where $F_{\mathbf{y}}^*$ is the solution of another optimisation problem

$$F_{\mathbf{y}}^*(\mathbf{x}, \xi) = \min_{\mathbf{y} \in \mathcal{Y}(\mathbf{x}, \xi)} F_{\mathbf{y}}(\mathbf{x}, \mathbf{y}, \xi) \quad (2.18)$$

with $\mathcal{Y}(\mathbf{x}, \xi)$ in general dependent on both $\mathbf{x}$ and ξ . When (2.18) has a unique minimiser, we denote it by $\mathbf{y}^* = \mathbf{y}^*(\mathbf{x}, \xi)$. Under this decomposition, the stochastic programming problem (2.15) becomes

$$\min_{\mathbf{x} \in \mathcal{X}} [F_{\mathbf{x}}(\mathbf{x}) + \mathbb{E}(F_{\mathbf{y}}^*(\mathbf{x}, \xi))] \quad (2.19)$$

and the sample average approximation (2.16) becomes

$$\min_{\mathbf{x} \in \mathcal{X}} \left[F_{\mathbf{x}}(\mathbf{x}) + \frac{1}{n} \sum_{i=1}^n F_{\mathbf{y}}^*(\mathbf{x}, \xi_i) \right]. \quad (2.20)$$

Such problems involve two optimisation problems, one nested within the other. Determining $\mathbf{x}^*$ is the *first stage* problem. Determining $F_{\mathbf{y}}^*(\mathbf{x}, \xi)$ given $\mathbf{x}$ and ξ is the *second stage* problem and is solved deterministically *given a particular value* of $\mathbf{x}$ and ξ , a property known as *recourse*. In contrast, the first stage problem for $\mathbf{x}^*$ is done *without* knowledge of the value of ξ ; picking $\mathbf{x}$ to minimise the expectation $\mathbb{E}(F(\mathbf{x}, \xi))$ is hence a natural choice.

2.7 Planning by stochastic programming

In this section, we demonstrate that power system planning models are sample average approximations to two-stage stochastic programming problems. In particular, the system design and operation are the first and second stage decisions, with uncertainty coming from the demand and weather data. The design $\mathbf{D}$ that min-

Two-stage stochastic programming	Power system planning
i (sample index)	t (time step)
n (sample size)	$ \mathcal{T} $ (number of time steps / periods)
$\mathbf{x}$ (stage 1 decision)	$\mathbf{D}$ (system design)
$\mathbf{y}$ (stage 2 decision)	$\mathbf{O}$ (system operation)
$\boldsymbol{\xi}$ (random variable)	$\boldsymbol{\xi}$ (demand and weather data)
$F_{\mathbf{x}}(\mathbf{x})$ (stage 1 cost)	$\text{cost}_{\text{install}}$ (install cost)
$F_{\mathbf{y}}^*(\mathbf{x}, \boldsymbol{\xi})$ (stage 2 cost)	$\text{cost}_{\text{operate}}^*(\mathbf{D}, \boldsymbol{\xi})$ (operation cost)

Table 2.1: Correspondence between two-stage stochastic programming (Section 2.6.2) and power system planning problems (Section 2.3.2).

imises the expected cost is the solution to the stochastic programming problem, which we approximate by $\hat{\mathbf{D}}$, the sample average approximation obtained from a planning model run on demand and weather data. In this section, we provide the mathematical argument and comment briefly on the implications and assumptions. For a more detailed discussion, see [Wallace and Fleten, 2003](#).

2.7.1 Mathematics

We can write the planning problem (2.10) as a two-stage sample average approximation of the form (2.20):

$$(2.10) = |\mathcal{T}| \times \min_{\mathbf{D} \geq \mathbf{0}} \left[\text{cost}_{\text{install}}(\mathbf{D}) + \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \text{cost}_{\text{operate}}^*(\mathbf{D}, \boldsymbol{\xi}_t) \right] \quad (2.21)$$

$$= n \times \min_{\mathbf{x} \in \mathcal{X}} \left[F_{\mathbf{x}}(\mathbf{x}) + \frac{1}{n} \sum_{i=1}^n F_{\mathbf{y}}^*(\mathbf{x}, \boldsymbol{\xi}_i) \right] = (2.20) \quad (2.22)$$

where the correspondence of terms is provided in Table 2.1. The random variable $\boldsymbol{\xi}$ represents the demand and weather. The first stage decision $\mathbf{D}$ is the system design, with install cost $\text{cost}_{\text{install}}(\mathbf{D})$. The second stage decision is the system operation $\mathbf{O}_t$, with

$$\text{cost}_{\text{operate}}^*(\mathbf{D}, \boldsymbol{\xi}_t) = \min_{\mathbf{O}_t} \text{cost}_{\text{operate}}(\mathbf{O}_t) \quad \text{subject to } (2.2)-(2.9) \quad (2.23)$$

being the cost to fulfill demand (the solution to the operation problem at time t). It depends on the system design $\mathbf{D}$ and demand/weather data $\boldsymbol{\xi}_t$ through operational constraints (2.2)-(2.9).

The term being minimised in (2.21) is the sample average *total* (sum of install and operation) cost per time step. This relies on $\text{cost}_{\text{install}}(\mathbf{D})$ being the install cost

per time step of plant lifetime and the subsequent normalisation of the operation costs by $\frac{1}{|\mathcal{T}|}$, where $|\mathcal{T}|$ is the number of time steps in $\mathcal{T}$.

2.7.2 Remarks

Intuitively, we seek the system design that minimises the total cost. This must be done under uncertainty; the exact demand and weather events that will occur are not known at the time of investment decisions. If we knew the underlying demand and weather distribution, the system design $\mathbf{D}$ that minimises the *expected* costs is a natural candidate. However, we do not; $\hat{\mathbf{D}}$, the optimum across a sample of demand and weather data $(\xi_t)_{t \in \mathcal{T}}$, is the only available quantity. This is determined using a power system planning model of the form of (2.10).

Statistical theory justifies the use of power system planning models with historic data. In particular, under certain conditions, the estimated optimal design is consistent, meaning $\hat{\mathbf{D}}((\xi_t)_{t \in \mathcal{T}}) \xrightarrow{\mathbb{P}} \mathbf{D}$ as the number of time steps $|\mathcal{T}| \rightarrow \infty$. As discussed in Appendix A.4.2, the key requirements are uniform convergence of the loss function (considered as a function of $\mathbf{D}$) and that the true optimum $\mathbf{D}$ does not lie on the border of the feasible region. These conditions are met for our problems except in the case where we do not allow unmet demand; in this case the optimal solution is usually one with zero unused generation capacity, on the border between the feasible and unfeasible region. In this setting, design estimates may still be consistent, but the proofs we cite do not apply directly.

The consistency conditions of Appendix A.4.2 also require that the ξ_t (demand and weather events) are independent and identically distributed (IID) and that the operation problem can be solved independently for each time step (a property known as *separability*, see Section 3.4.1.1). This is usually not the case; the operation problem may have constraints linking time steps (such as limits on the rate of change of generation levels or storage level continuity) and time series may have temporal dependence structures. However, judiciously sampling longer blocks (e.g. weeks from across the year) can generate realistic samples; see Section 6.3.1 for details.

The operation problem is solved with *recourse*: generation, transmission and storage decisions are determined *after* demand and weather patterns are known. We hence assume perfect foresight at operational time scales. In real-world power systems, many operational decisions are made one day ahead using demand and weather forecasts. With long-term storage, perfect foresight is less realistic — in this case, storage (dis)charge decisions are made with perfect knowledge of the future.

3. Power system models for time series analysis

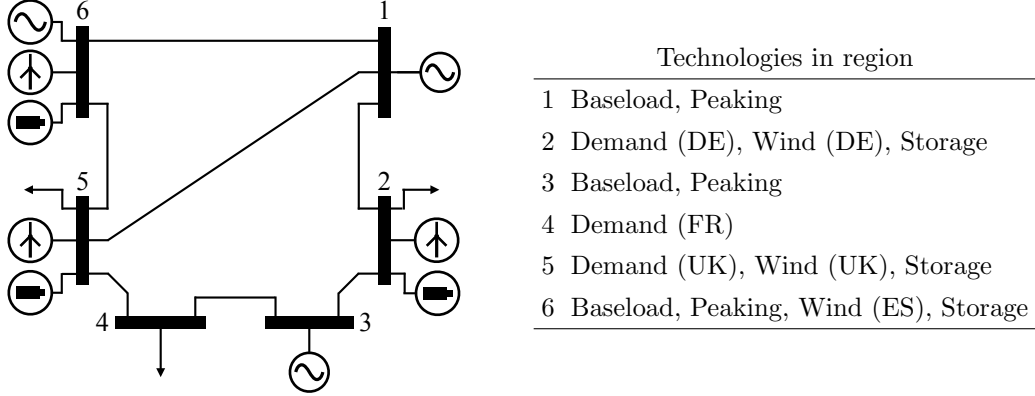


Figure 3.1: Model configuration. Demand, generation (baseload, peaking, wind) and storage are distributed across six regions connected by seven transmission lines. Regions 2, 4, 5 and 6 use time series data from Germany (DE), France (FR), United Kingdom (UK) and Spain (ES) respectively.

In this section, we introduce a set of power system models designed for the study of time series uncertainty, which we use in our case studies in Chapters 4 and 6. We make these models publicly available for future research as an open-source software package (see Section 3.2). Tutorials with these models have been used as introductions to energy system modelling.

This chapter is structured as follows. Section 3.1 provides a high-level overview of the models, with the basic information required for the remainder of the thesis. 3.2 discusses the functionality that our open-source software package provides. Section 3.3 examines how the optimisation problems' computational costs depend on simulation length, constraints and solvers. Section 3.4 introduces a number of decomposition methods to simplify optimisation problems. Section 3.5 provides mathematical details, including costs, time series inputs and mathematical details.

3.1 Overview

The power system planning problem in our case studies is a particular case of that introduced in Section 2.3. Figure 3.1 shows its topology: demand, generation and storage technologies distributed across six regions with seven transmission lines. It

is based on a renewable version of the *IEEE six-bus* system introduced by Kamalinia and Shahidehpour [2010] and Kamalinia et al. [2011]. The model takes six input time series of hourly demand levels (in Regions 2, 4 and 5) and wind generation potentials (in Regions 2, 5 and 6).

We consider three generation technologies. *Baseload* (b) and *peaking* (p) are weather-independent and differ only in their install and generation costs; baseload has a higher install cost but lower generation cost. *Wind* (w) generation is free but constrained by weather-dependent *generation potentials*: generation as a fraction of its theoretical maximum under perfect wind conditions. *Unmet demand* (u) is considered a fourth generation technology with no install cost but a very high generation cost, corresponding to the *value of lost load*; it allows an economic trade-off between unserved energy and generation capacity which is almost never used. We can disallow it, in which case 100% of demand must be met. See Table 3.1 for details on the generation technologies.

The system contains a single storage technology, allowing energy to be saved for later use. It represents long-term storage, for which discharge rates are small compared to total capacity (e.g. large hydro dams). For such technologies, costs depend primarily on *energy* (in MWh) and not *power* (instantaneous (dis)charge rate, in MW) capacities. For this reason, we only price energy capacity in the cost function. Furthermore, power constraints, which cap discharge rates at installed power capacity, are rarely enforced since these discharge rates are usually much smaller than capacity. Hence, we relax (ignore) power constraints without significant impact on the problem's solution.

Four vectors define the state of the system at time t (with terms as defined in Section 2.3 and Nomenclature):

- design $\mathbf{D} = [\text{cap}_{i,r}^{\text{gen}}, \text{cap}_{r,r'}^{\text{tr}}, \text{cap}_r^{\text{sto}} \mid i \in \mathcal{I}; r, r' \in \mathcal{R}]$: installed capacities of generation, transmission and storage technologies, in MW, MW and MWh respectively
- operation $\mathbf{O}_t = [\text{gen}_{i,r,t}, \text{tr}_{r,r',t}, \text{ch}_{r,t} \mid i \in \mathcal{I}; r, r' \in \mathcal{R}]$: generation, transmission and storage (dis)charge rates, all in MW
- storage energy levels $\mathbf{sto}_t = [\text{sto}_{r,t} \mid r \in \mathcal{R}]$, in MWh
- demand and weather data $\boldsymbol{\xi}_t = [d_{r,t}, \lambda_{w,r,t} \mid r \in \mathcal{R}]$: demand levels and wind generation potentials, in MW and dimensionless respectively

We run three planning problems with different constraints. The first is the *continuous linear program (CLP)* problem, in which all generation and transmission technologies may be installed to any nonnegative capacity but there is no storage capacity. The *mixed integer linear program (MILP)* problem is identical (still no storage) except two additional constraints on baseload (both of which do *not* apply to peaking, wind or storage): a *ramping* constraint that generation levels can change at most 20% of capacity per hour, and a restriction of installed capacity to units of 3GW, roughly a large nuclear plant. The last problem is the *continuous linear program with storage (CLPwS)* problem, which is the *CLP* problem with the ability to build storage.

3.2 Open-source software package

We make our power system models publicly available by uploading the model-defining files and time series data (Section 3.5.2) to an online repository with a permissive license. They are free to use for the energy and other research communities, and a tutorial based on them have served as introductions to energy system modelling. The data, code and tutorial can be found at doi:[10.5281/zenodo.7025098](https://doi.org/10.5281/zenodo.7025098).

Users can run our models via a `python` package which allows the input time series to be changed easily. Essentially, we provide a programmatic implementation of (2.13) and (2.14), where power system models are interpreted as functions from time series data to cost-optimal design or operational variables. Constraints, such as storage, integer generation capacities or whether to allow unserved energy, can be enabled or disabled via function arguments.

Our code works as follows. Via function arguments, a user specifies the (possibly subsampled) time series data and optional constraints. Our package creates the model using the `calliope` model-generation package (Pfenninger and Pickering, 2018) and solves it using a solver (see next section for details). Our code then allows the user to easily inspect model results.

The advantages of our code are that (1) the model’s technologies and topology are available off-the-shelf and do not have to be specified, (2) constraints and settings can be adjusted easily as function arguments, and (3) both time series and results are easily accessible — the full model can be created, solved and examined in a single `python` session with a few lines of code, without having to read/write results from/to saved files.

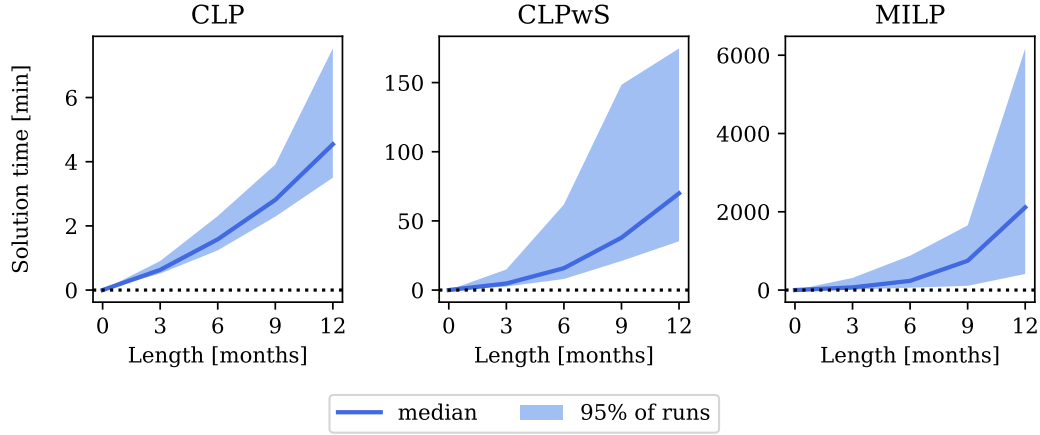


Figure 3.2: Distribution (across 40 runs) of solution times as a function of simulation length. Plots have different y -scales.

3.3 Solvers and computational cost

The mathematical nature of the three problems impacts their computational complexities. Figure 3.2 shows the distribution of solution times as a function of simulation length (number of time steps) for each planning problem (*CLP*, *CLPwS* and *MILP*). We solve them using the *cbc* solver [Forrest et al., 2020](#) on a 2.7GHz Intel Core i5-5257U processor with 8 GB of RAM.

The *CLP* problem has the lowest computational complexity; a full-year simulation takes 3-7 minutes, and solution times increase slightly faster than linearly. The inclusion of storage in the *CLPwS* model, which links constraints in different time steps, makes solution times scale more quickly and with a larger variability — 95% of full-year simulations require 40-160 minutes of computing time. The impact of integer variables in the *MILP* model amplifies this further: solution times scale exponentially and a full-year simulation can take anywhere from six hours to four days. For all models, memory (RAM) requirements scale linearly, at about 120MB per month of hourly time steps.

We use the *cbc* solver because it may be used under a free license (anyone that uses our software package can also solve the optimisation problems) and is easily implementable on computing clusters (allowing many jobs in parallel). We also experimented with the commercial solver *gurobi*. It is around 30% faster for the *CLP* and *CLPwS* models, and significantly faster for the *MILP* model (95% of full-year simulations took 1.5-2.5 hours instead of 6-100). However, the scaling with respect to simulation length (doubling with every additional 3 months) is unchanged.

3.4 Decomposition methods

Some simple classes of power system planning problems can be decomposed or solved via simple techniques. This is usually not the case for the more complicated models in this thesis, but such approaches can serve as useful heuristics which inform other solution strategies, including our own. These methods are not our original work; we cite previous texts where appropriate.

3.4.1 Operation problem

For the operation problem, we determine system operation $(\mathbf{O}_t)_{t \in \mathcal{T}}$ (generation, transmission and storage decisions in each time step t) with a fixed system design (installed capacities).

3.4.1.1 Separable structure

When each constraint depends on exactly one time step t , the operation problem is *separable* [Boyd and Vandenberghe, 2009, chapter 10] and may be solved in each time step independently and in parallel. The problem's separability disappears with *linking* constraints that relate system operation in different time steps; examples include *ramping* (limits on rates of change of generation levels) or electricity storage.

3.4.1.2 Merit order stacking

If the operation problem is separable and contains only one region, we determine generation levels by *merit-order stacking* of technologies in ascending order of generation cost, up to their installed capacities, until demand is met (Figure 3.3). While not possible for models with multiple regions or a non-separable structure, this method often gives approximate answers that are intuitively helpful.

3.4.1.3 Horizons and windows

We may (approximately) solve operation problems by decomposing them into sequential *horizons* and *windows* as shown in Figure 3.4. Instead of optimising for each temporal period at once, we solve only the first *horizon* (e.g. 48 hours) and store a subset of this, called the *window* (e.g. 24 hours). We then move forward by the window length and repeat, using the state at the end of the previous window to satisfy any linking constraints.

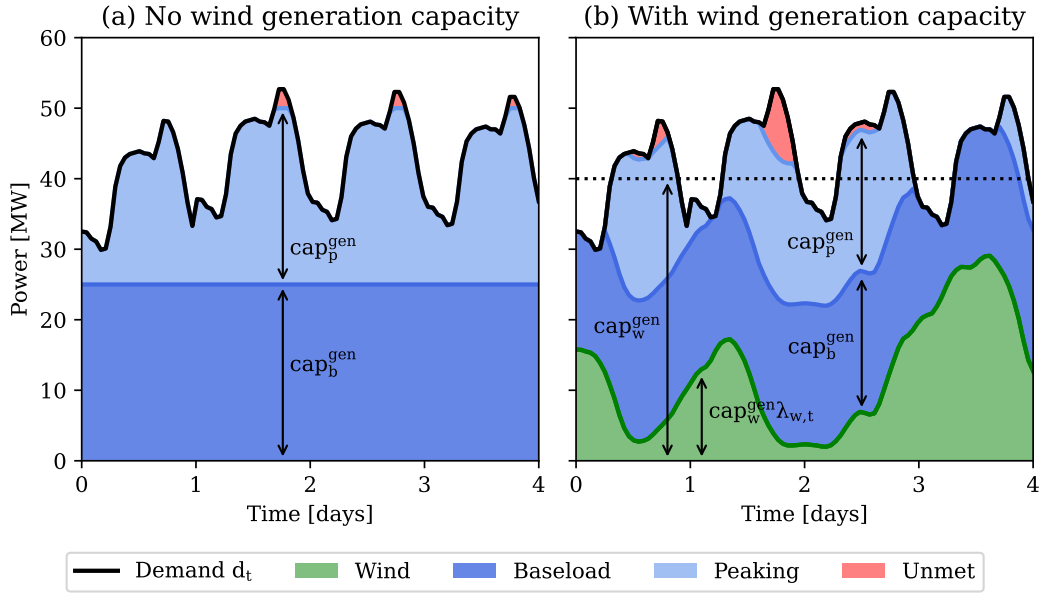


Figure 3.3: *Merit-order* stacking of generation in ascending order of generation cost. Wind is run first, up to its installed capacity times wind generation potential $\lambda_{w,t}$, followed by baseload and peaking up to their installed capacities.

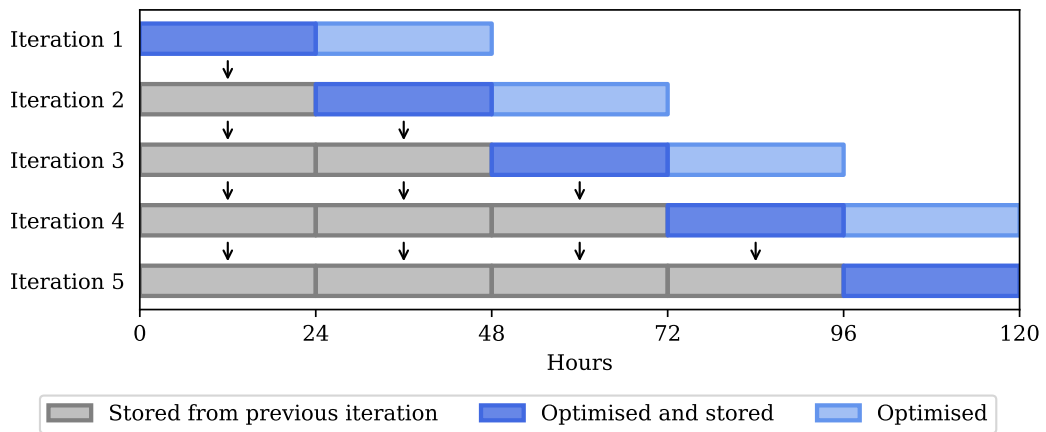


Figure 3.4: Approximately solving a 120-hour operation problem via a 48-hour horizon and 24-hour window.

We generate feasible solutions by enforcing all constraints within a window as well as linking constraints with previous ones. Furthermore, by considering extra time beyond the immediate window (the horizon), the optimisation is performed with some knowledge of the future. With a horizon length long enough to cover temporal dependencies, solutions are close to optimal. For our models, we find that solutions are exact when using horizons and windows at least four and two weeks respectively (without storage) and one year and six months respectively (with storage).

3.4.2 Planning problem

The planning problem is mathematically different from the operation problem since the system design $\mathbf{D}$ (installed generation, transmission and storage capacities) is a decision variable instead of being fixed. The components of $\mathbf{D}$ are called *complicating variables* [Cao et al., 2019] since they appear in constraints at *every* time step, see (2.6)-(2.8). This makes the problem non-separable and prevents its decomposition into temporal blocks as we do for the operation problem. Hence, determining $\hat{\mathbf{D}}$ is more computationally involved than determining the operational variables; this structure motivates solution methods that we describe below.

3.4.2.1 Load duration curve

When the operational subproblem admits merit-order stacking, we can solve the planning problem graphically using the *load duration curve* (LDC) as described by [Stoft 2002]. Given an x -value, the LDC's y -value is the demand level exceeded a proportion x of the year. Practically, it is the demand levels plotted in descending order, as in Figure 3.5.

With renewables, we consider the *net demand* d_t^{net} that must be met with dispatchable (non-renewable) generation. For example, for a system with wind power generation,

$$d_t^{\text{net}} := d_t - \text{gen}_{w,t} = \begin{cases} d_t - \text{cap}_w^{\text{gen}} \lambda_{w,t} & \text{if } d_t \geq \text{cap}_w^{\text{gen}} \lambda_{w,t} \\ 0 & \text{otherwise.} \end{cases} \quad (3.1)$$

We create the *net load duration curve* from the net demand in the same way as the LDC from the demand. Each value of installed wind capacity $\text{cap}_w^{\text{gen}}$ leads to a different net LDC as illustrated in Figure 3.5. The regular LDC is the net LDC

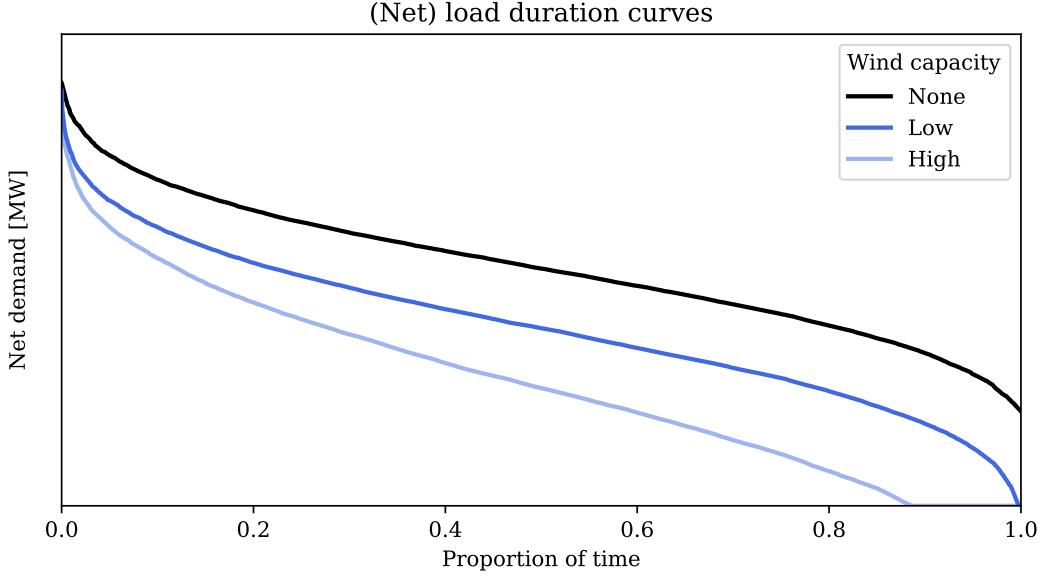


Figure 3.5: Net *load duration curves* (LDCs) with varying wind generation capacity. For a given x -value, these show the net demand (demand minus available wind generation) exceeded a proportion x of the time.

with a wind capacity $\text{cap}_w^{\text{gen}}$ of zero.

We can graphically solve planning problems by combining the (net) LDC with *screening curves*, which show the sum of install and generation costs as a function of the proportion of time a technology generates. With constant generation costs, screening curves are lines with install cost (per time step) as y -intercept and generation cost as slope. We produce screening curves only for dispatchable technologies and not variable renewables such as solar and wind.

Screening curves for hypothetical *baseload* and *peaking* technologies, as well as unmet demand, are shown in Figure 3.6(a). The cheapest generation capacity depends on the proportion of time it is run for. *Baseload*, with a high install but low generation cost, is the cheapest technology when run at least 76% of the time. *Peaking* is cheapest for any demand levels that occur between 5% and 76% of the time. It is not cost-effective to build generation capacity for any demand that occurs less than 5% of the time, as the cost of installing rarely-used generation capacity is lower than the value of lost load. The points at which the cheapest technology changes are indicated by vertical dotted lines. The hypothetical technologies used in this example are not those used in our planning models.

Combining screening curves with load duration curves determines cost-optimal installed capacities of conventional generation technologies as shown in Figure 3.6(b).

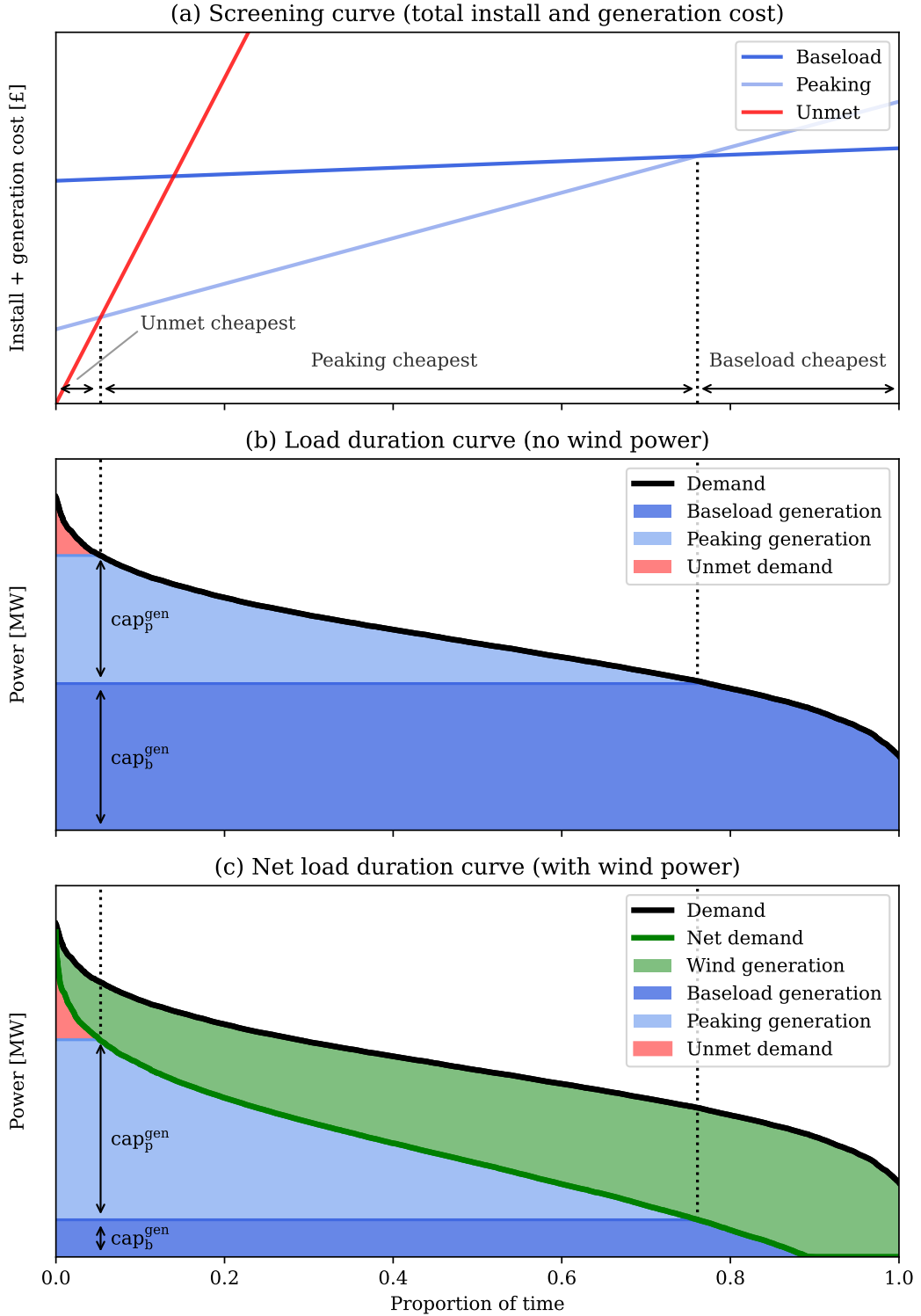


Figure 3.6: Cost-optimal generation capacities of dispatchable (weather-independent) technologies. A screening curve (a) is the total cost as a function of the proportion of time generating. The points at which the cheapest technology changes determine the cost-optimal generation capacity by comparison with the load duration curve (b) or the net load duration curve (c).

For example, the optimal baseload capacity is the demand level that occurs at least 76% of the time, corresponding to $x = 0.76$ on the LDC. A similar process is used to obtain the peaking capacity. Any demand above baseload plus peaking capacity is unmet.

When variable renewables such as solar and wind are allowed, this method determines the optimal installed capacities of conventional technologies *conditional* on the renewable capacity [Kennedy, 2005, De Jonghe et al., 2011]. This is shown in Figure 3.6(c). In this example, the optimal wind capacity must still be determined. However, the method of (net) LDCs reduces the problem’s dimension, since the optimisation must be performed over one variable (wind generation capacity) instead of three (wind, baseload and peaking capacity). If e.g. solar were also considered, the optimisation problem must be solved across wind and solar capacities; the capacities of conventional technologies conditioned on these is determined via load duration curves.

3.5 Mathematical details

In this section, we discuss technologies (Section 3.5.1), time series (Section 3.5.2) and the precise optimisation problem (Section 3.5.3) for our planning model. For parameter values, see Table 3.1.

3.5.1 Technologies

3.5.1.1 Generation

Peaking is based on a UK CCGT H class reactor (commonly used as load-matching generation) described by [Department for Business Energy and Industrial Strategy, 2016, Chart 14], with build cost £100/kW per year of plant lifetime. We assume a gas price of £20 per MWh¹ and generation efficiency of 55%, meaning $\frac{20}{0.55} \approx £35/\text{MWh}$, and carbon emissions of 400 grams of CO₂ equivalent per KWh [Staffell]. *Baseload* is a technology with a higher install cost (£300/kW per year) but lower generation cost, roughly based on new nuclear capacity. *Wind* power is a weighted average of onshore and offshore from [Department for Business Energy and Industrial Strategy, 2016]. *Unmet demand* is a generation technology with no installation cost but a generation cost of £6,000/MWh, a UK value of lost load

¹This value was roughly correct at the start of this thesis in 2018. At the time of submission (2022), European gas prices were around 15 times higher.

Technology (i)	Monetary cost			Emissions [kgCO ₂ e/MWh]	Storage	
	Install [£/MWyr]	Install [£/MWhyr]	Generation [£/MWh]		Efficiency [1]	Self-loss [1/hr]
<i>Generation</i>	C_i^{gen}		F_i	E_i		
Baseload (b)	300,000	—	5	200	—	—
Peaking (p)	100,000	—	35	400	—	—
Wind (w)	100,000	—	—	—	—	—
Unmet (u)	—	—	6,000	—	—	—
<i>Transmission</i>	$C_{r,r'}^{\text{tr}}$					
Region 1-5	150,000	—	—	—	—	—
Other	100,000	—	—	—	—	—
<i>Storage</i>		C^{sto}			e^{sto}	l^{sto}
Storage	—	1,000	—	—	0.95	0.00001

Table 3.1: Technologies. Install costs are expressed per year of infrastructure lifetime. Carbon emissions are expressed in kg CO₂ equivalent warming potential. Storage efficiency is dimensionless.

estimate [\[Ellexon, 2018\]](#).

3.5.1.2 Transmission

The costs of transmission technologies are chosen to balance generation install costs. If transmission install costs are too high, generation occurs locally and the model separates into disjoint power systems. If they are too low, the model approximates a single region. We choose values in the middle: £150,000 per MW per year of plant lifetime for transmission from Region 1 to 5 and £100,000 for all others.

3.5.1.3 Storage

We consider a single storage technology, with parameter values chosen to represent multi-day storage potential. This technology has a high efficiency (95%, corresponding to a 90.25% round-trip efficiency). The install cost is chosen so that the optimal design and operation contains inter-day storage. Since we model a long-term storage technology, we price and constrain only storage's *energy* capacity (MWh), and not its *power* capacity (MW).

3.5.2 Time series

The model takes six input time series in total: demand in Regions 2, 4 and 5; and wind *generation potentials* (generation as proportion of maximum) in Regions 2, 5 and 6. They come from a dataset of European countries over the period 1980-

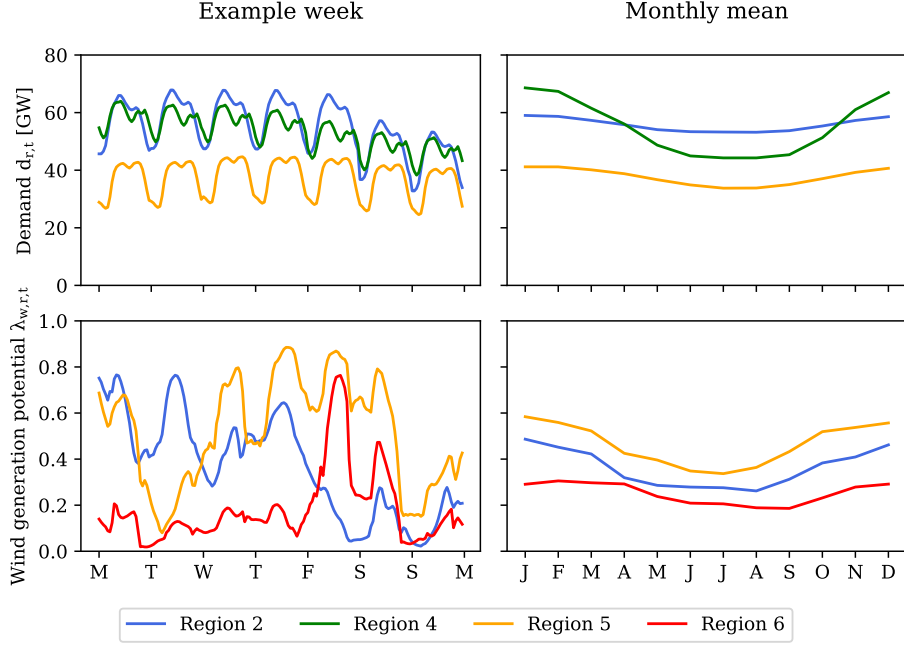


Figure 3.7: Demand and weather (wind generation potential) time series.

2017, introduced by [Bloomfield et al., 2019](#) and available at [Bloomfield et al., 2020](#). They are based on national electricity demand and weather time series from MERRA reanalysis [Rienecker et al., 2011](#).

Figure [3.7](#) shows each time series across an example week as well as their monthly average values. For demand, the annual, weekly and annual cycles are clearly visible. The strength of these cycles depends on the country; for example, demand in Region 4 (data from France) has a stronger annual cycle due to higher levels of electrified heating. The wind time series have an annual and daily cycle but no weekly trend.

3.5.3 Mathematical formulation

In this section, we provide the precise mathematical optimisation problem associated with our planning model. This is to minimise

$$\sum_{r \in \mathcal{R}} \left[\frac{T}{8760} \left(\underbrace{\sum_{i \in \mathcal{I}} C_i^{\text{gen}} \text{cap}_{i,r}^{\text{gen}}}_{\text{install cost, generation}} + \underbrace{\sum_{r' \in \mathcal{R}} C_{r,r'}^{\text{tr}} \text{cap}_{r,r'}^{\text{tr}}}_{\text{install cost, transmission}} + \underbrace{C_r^{\text{sto}} \text{cap}_r^{\text{sto}}}_{\text{install cost, storage}} \right) + \underbrace{\sum_{t \in \mathcal{T}} \sum_{i \in \mathcal{I}} F_i^{\text{gen}} \text{gen}_{i,r,t}}_{\text{generation cost}} \right] \quad (3.2)$$

by optimising over decision variables $\mathbf{D}$ and $(\mathbf{O}_t)_{t \in \mathcal{T}}$, where

$$\mathbf{D} = [\text{cap}_{i,r}^{\text{gen}}, \text{cap}_{r,r'}^{\text{tr}}, \text{cap}_r^{\text{sto}} \mid i \in \mathcal{I}; r \in \mathcal{R}; r' \in \mathcal{R}] \quad (3.3)$$

$$\mathbf{O}_t = [\text{gen}_{i,r,t}, \text{tr}_{r,r',t}, \text{ch}_{r,t} \mid i \in \mathcal{I}; r \in \mathcal{R}; r' \in \mathcal{R}] \quad (3.4)$$

subject to

$$\text{cap}_{b,r}^{\text{gen}} \big|_{r \notin \{1,3,6\}} = \text{cap}_{p,r}^{\text{gen}} \big|_{r \notin \{1,3,6\}} = \text{cap}_{w,r}^{\text{gen}} \big|_{r \notin \{2,5,6\}} = 0 \quad (3.5)$$

$$\text{cap}_{r,r'}^{\text{tr}} \big|_{(r,r') \notin \{(1,2), (1,5), (1,6), (2,3), (3,4), (4,5), (5,6)\}} = 0 \quad (3.6)$$

$$\text{cap}_r^{\text{sto}} \big|_{r \notin \{2,5,6\}} = 0 \quad (3.7)$$

$$\sum_{i \in \mathcal{I}} \text{gen}_{i,r,t} + \sum_{r' \in \mathcal{R}} \text{tr}_{r',r,t} = d_{r,t} + \text{ch}_{r,t} \quad \text{for all } r, t \quad (3.8)$$

$$\text{tr}_{r,r',t} + \text{tr}_{r',r,t} = 0 \quad \text{for all } r, r', t \quad (3.9)$$

$$\text{sto}_{r,0} = 0 \quad \text{for all } r \quad (3.10)$$

$$\text{sto}_{r,t+1} = (1 - l^{\text{sto}}) \text{sto}_{r,t} + \begin{cases} e^{\text{sto}} \text{ch}_{r,t} & \text{if } \text{ch}_{r,t} \geq 0 \\ \frac{1}{e^{\text{sto}}} \text{ch}_{r,t} & \text{if } \text{ch}_{r,t} < 0 \end{cases} \quad \text{for all } r, t \quad (3.11)$$

$$0 \leq \text{gen}_{i,r,t} \leq \text{cap}_{i,r}^{\text{gen}} \quad \text{for all } i, r, t \quad (3.12)$$

$$0 \leq \text{gen}_{w,r,t} \leq \text{cap}_{w,r}^{\text{gen}} \lambda_{w,r,t} \quad \text{for all } r, t \quad (3.13)$$

$$|\text{tr}_{r,r',t}| \leq \text{cap}_{r,r'}^{\text{tr}} + \text{cap}_{r',r}^{\text{tr}} \quad \text{for all } r, r', t \quad (3.14)$$

$$0 \leq \text{sto}_{r,t} \leq \text{cap}_r^{\text{sto}} \quad \text{for all } r, t \quad (3.15)$$

$$[\text{MILP only}] \quad |\text{gen}_{b,r,t} - \text{gen}_{b,r,t+1}| \leq 0.2 \text{cap}_{b,r}^{\text{gen}} \quad \text{for all } r, t \quad (3.16)$$

$$[\text{MILP only}] \quad \text{cap}_{b,r}^{\text{gen}} \in 3\mathbb{Z} \quad \text{for all } r \quad (3.17)$$

$$[\text{CLP and MILP only}] \quad \text{cap}_r^{\text{sto}} = 0 \quad \text{for all } r \quad (3.18)$$

$$[\text{optional}] \quad \text{gen}_{u,r,t} = 0 \quad \text{for all } r, t. \quad (3.19)$$

For definitions of terms and parameter values, see [Nomenclature](#) and Table [3.1](#) respectively. To avoid solution nonuniqueness, costs are perturbed slightly ($< 0.1\%$) in different regions. The vectors $\mathbf{D}$ and $(\mathbf{O}_t)_{t \in \mathcal{T}}$ are the design and operational decision variables. The factor $\frac{T}{8760}$ normalises install costs to the same temporal scale as generation costs, since C_i^{gen} and $C_{r,r'}^{\text{tr}}$ are costs per year of plant lifetime and there are 8760 hours (time steps) in a year.

We have the following constraints. [\(3.5\)](#)-[\(3.7\)](#) specify the generation, transmission and storage topology. [\(3.8\)](#) indicates that generation plus transmission into a

region equals demand plus storage charging. (3.9) is the transmission balance. (3.10) and (3.11) specify empty initial storage levels and how they change in time. (3.12)-(3.13) ensure generation does not exceed installed capacity (for thermal technologies) or installed capacity times generation potential (for wind). (3.14) limits transmitted power to installed capacity. (3.15) constrains storage levels to lie within energy bounds; we do not constrain instantaneous storage power. (3.16) (the baseload ramping constraint) and (3.17) (baseload is built in units of 3GW) are active only for the *MILP* problem. (3.18) removes storage in the *CLP* and *MILP* models; we consider storage only in the *CLPwS* problem. (3.19) optionally stipulates that all demand is met by generation (no unmet demand).

4. Efficient design estimates without storage

In this chapter, we introduce a subsampling algorithm to improve optimal design estimates from power system planning models. It allows users to estimate, at significantly reduced computational cost, the optimal system design across a long sample of demand and weather data. This reduces estimates' bias and variance under demand and weather uncertainty. The algorithm is published in the following papers, where the second generalises the first:

- [Hilbers et al., 2019] AP Hilbers, DJ Brayshaw, A Gandy (2019). Importance subsampling: improving power system planning under climate-based uncertainty. *Applied Energy*.
- [Hilbers et al., 2020] AP Hilbers, DJ Brayshaw, A Gandy (2020). Importance subsampling for power system planning under multi-year demand and weather uncertainty. In *2020 IEEE International Conference on Probabilistic Methods Applied to Power Systems*.

Models, data, methods and code are available at doi:[10.5281/zenodo.3841154](https://doi.org/10.5281/zenodo.3841154). The algorithm is applicable to models *without* constraints linking time steps; See Chapter 5 for an extension to inter-day storage and linking constraints.

This chapter is structured as follows. Section 4.1 provides background on the problem's mathematical details and established approaches. Section 4.2 describes our contribution. Section 4.3 introduces the method and Section 4.4 provides a case study of its application on the test planning models of Chapter 3. Section 4.5 makes concluding remarks.

4.1 Background

4.1.1 Context: time series aggregation

This chapter introduces a *time series aggregation* scheme for planning problems as introduced in Section 2.3.2. We seek $\mathbf{D}$, the system design that minimises the sum of install and expected operation costs under demand and weather uncertainty. Without access to the demand and weather distribution, we instead calculate

$\hat{\mathbf{D}}_{\mathcal{T}}$, the minimiser across the longest available sample of demand and weather data $(\boldsymbol{\xi}_t)_{t \in \mathcal{T}}$. This is (in principle) calculated using a power system planning model:

$$\hat{\mathbf{D}}_{\mathcal{T}} = \text{PSM}_{\text{plan}}((\boldsymbol{\xi}_t)_{t \in \mathcal{T}}). \quad (4.1)$$

Due to computational cost (Section 2.4), calculating $\hat{\mathbf{D}}_{\mathcal{T}}$ directly may be unfeasible. Instead, we may run planning models on a shorter (possibly weighted) time series to give an estimate $\hat{\mathbf{D}}_{\mathcal{A}}$ at reduced computational cost. The process of creating a shorter time series is called *time series aggregation* and induces approximation errors. For example, choosing a single year or downsampling to lower temporal resolution are shown to induce significant errors in Sections 2.4–2.5. In this chapter, we introduce a method with lower approximation errors than established approaches. First, however, we present the mathematics of time series aggregation and how it reduces computational cost.

4.1.2 Mathematical background

4.1.2.1 Time series aggregation

Consider a time series of demand and weather data $(\boldsymbol{\xi}_t)_{t \in \mathcal{T}}$. Partition it into *periods* — usually days or weeks. Let $\mathcal{P}$ be the set of periods in $\mathcal{T}$, and let $\boldsymbol{\xi}_p^{\text{P}}$ contain data in period p . For notational simplicity, we drop the superscript P and differentiate by index; $\boldsymbol{\xi}_t$ and $\boldsymbol{\xi}_p$ refer to data in time steps and periods respectively. For example, for daily periods with hourly time steps,

$$\boldsymbol{\xi}_p = [d_{p,1}, \dots, d_{p,24}, w_{p,1}, \dots, w_{p,24}] \quad (4.2)$$

where $d_{p,q}$ and $w_{p,q}$ are the demand and wind generation potential in hour q of period p . The time series $(\boldsymbol{\xi}_p)_{p \in \mathcal{P}}$ contains the same information as $(\boldsymbol{\xi}_t)_{t \in \mathcal{T}}$ but with vectors covering periods instead of individual time steps.

Time series aggregation maps $(\boldsymbol{\xi}_p)_{p \in \mathcal{P}}$ to a smaller set of *representative periods* $(\hat{\boldsymbol{\xi}}_{\mathcal{A}(p)})_{p \in \mathcal{P}}$. For example, in Figure 4.1:

$$\begin{aligned} (\boldsymbol{\xi}_p)_{p \in \mathcal{P}} &= (\boldsymbol{\xi}_1 \quad \boldsymbol{\xi}_2 \quad \boldsymbol{\xi}_3 \quad \boldsymbol{\xi}_4 \quad \boldsymbol{\xi}_5 \quad \boldsymbol{\xi}_6) \\ &\quad \downarrow \\ (\hat{\boldsymbol{\xi}}_{\mathcal{A}(p)})_{p \in \mathcal{P}} &= (\hat{\boldsymbol{\xi}}_1 \quad \hat{\boldsymbol{\xi}}_2 \quad \hat{\boldsymbol{\xi}}_1 \quad \hat{\boldsymbol{\xi}}_3 \quad \hat{\boldsymbol{\xi}}_2 \quad \hat{\boldsymbol{\xi}}_2). \end{aligned} \quad (4.3)$$

The mapping $\mathcal{A} : \mathcal{P} \rightarrow \mathcal{P}$ and representative periods $\hat{\boldsymbol{\xi}}_{\mathcal{A}(p)}$ can be determined in

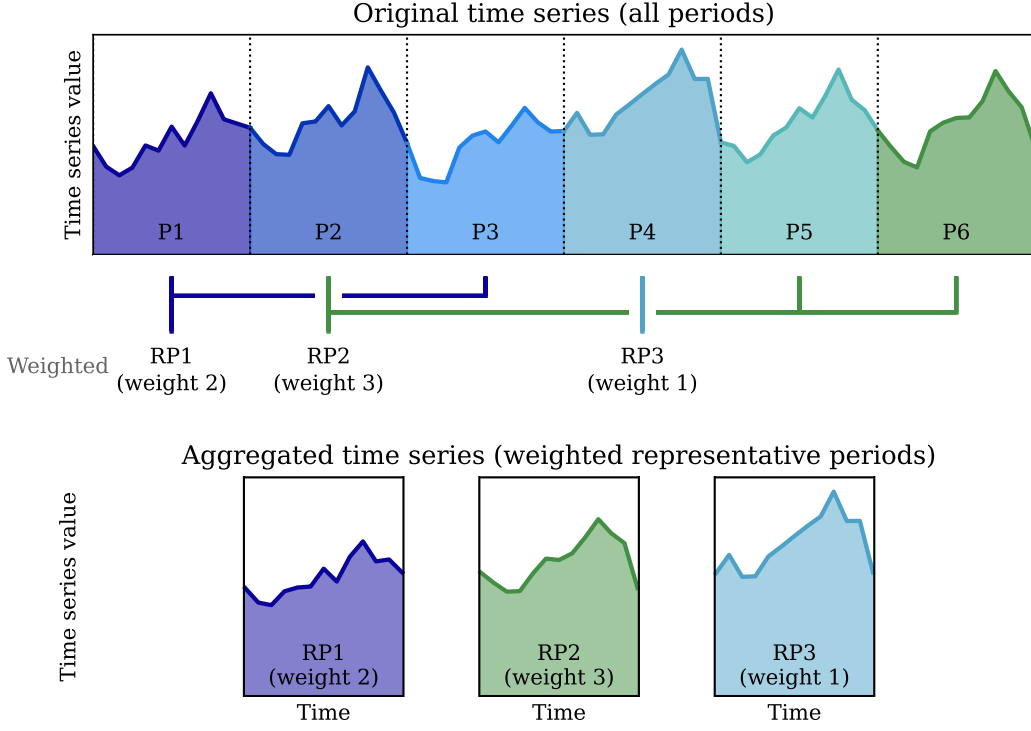


Figure 4.1: Time series aggregation of six periods (P1-6) to three weighted *representative periods* RP1-3.

various ways as discussed in Section 4.1.3.1. The design estimate $\hat{\mathbf{D}}_{\mathcal{A}}$ is subsequently determined from a power system planning problem run on the aggregated time series:

$$\hat{\mathbf{D}}_{\mathcal{A}} = \text{PSM}_{\text{plan}}((\hat{\xi}_{\mathcal{A}(p)})_{p \in \mathcal{P}}). \quad (4.4)$$

We show in Section 4.1.2.2 that, in the absence of constraints *linking* periods, design estimates $\hat{\mathbf{D}}_{\mathcal{A}}$ are independent of their order. In this case, we need only consider their relative frequencies, as illustrated in Figure 4.1:

$$(\hat{\xi}_{\mathcal{A}(p)})_{p \in \mathcal{P}} = (\hat{\xi}_1 \quad \hat{\xi}_2 \quad \hat{\xi}_1 \quad \hat{\xi}_3 \quad \hat{\xi}_2 \quad \hat{\xi}_2) \quad (4.5)$$

$$\stackrel{\text{equivalent}}{=} (\hat{\xi}_1 \text{ (weight 2)} \quad \hat{\xi}_2 \text{ (weight 3)} \quad \hat{\xi}_3 \text{ (weight 1)}) \quad (4.6)$$

where *equivalent* indicates an identical design estimate $\hat{\mathbf{D}}_{\mathcal{A}}$ when used in a planning model. We show in the next section that the planning problem requires less computational cost whenever $N_{\mathcal{P}_r}$, the number of unique representative periods $\{\mathcal{A}(p) \mid p \in \mathcal{P}\}$, is less than $N_{\mathcal{P}}$, the number of periods in the full time series $\{p \mid p \in \mathcal{P}\}$.

4.1.2.2 Impact on planning problem

In this section we explain how time series aggregation reduces computational cost. Consider a time series split into $N_{\mathcal{P}}$ periods with $N_{\mathcal{Q}}$ time steps each, and let (p, q) index time step q of period p . Consider also a toy power system planning model with one region and a single dispatchable generation technology with install cost C^{gen} (£/MW per time step of lifetime), generation cost F^g (£/MWh), generation capacity cap^{gen} (MW) and generation levels $\text{gen}_{p,q}$ (MWh). The demand levels $d_{p,q}$ (MWh) are the only time series input, so that $\boldsymbol{\xi}_p = [d_{p,1}, \dots, d_{p,N_{\mathcal{Q}}}]$ represents the vector of inputs per period. We use this simple toy planning model to illustrate time series aggregation without unnecessary technical detail; the same ideas apply to more complicated models also.

The planning problem is

$$\min \left[\underbrace{N_{\mathcal{P}} N_{\mathcal{Q}} C^{\text{gen}} \text{cap}^{\text{gen}}}_{\text{install cost}} + \underbrace{F^g \sum_{p=1}^{N_{\mathcal{P}}} \sum_{q=1}^{N_{\mathcal{Q}}} \text{gen}_{p,q}}_{\text{generation cost}} \right] \quad (4.7)$$

by optimising over $[\text{cap}^{\text{gen}}, \text{gen}_{p,q} \mid p=1 \dots N_{\mathcal{P}}, q=1 \dots N_{\mathcal{Q}}]$, subject to

$$\text{gen}_{p,q} = d_{p,q} \quad \text{for all } p = 1 \dots N_{\mathcal{P}}, q = 1 \dots N_{\mathcal{Q}} \quad (4.8)$$

$$\text{gen}_{pq} \in [0, \text{cap}^{\text{gen}}] \quad \text{for all } p = 1 \dots N_{\mathcal{P}}, q = 1 \dots N_{\mathcal{Q}}. \quad (4.9)$$

The constraints specify that generation equals demand (4.8) and remains within generation capacity (4.9). For this simple model, the solution is $\text{cap}^{\text{gen}} = \max(d_{p,q} \mid p = 1 \dots N_{\mathcal{P}}, q = 1 \dots N_{\mathcal{Q}})$ and $\text{gen}_{p,q} = d_{p,q}$, but for sake of illustration we apply time series aggregation.

Suppose we aggregate the input time series (demand levels) to a smaller number of representative periods:

$$d_{p,q} \mapsto \hat{d}_{\mathcal{A}(p),q} \quad \text{for all } p = 1 \dots N_{\mathcal{P}}, q = 1 \dots N_{\mathcal{Q}} \quad (4.10)$$

with $N_{\mathcal{P}_r}$ unique representative periods (unique values of $\mathcal{A}(p)$). The problem becomes:

$$\min \left[N_{\mathcal{P}} N_{\mathcal{Q}} C^{\text{gen}} \text{cap}^{\text{gen}} + F^g \sum_{p=1}^{N_{\mathcal{P}}} \sum_{q=1}^{N_{\mathcal{Q}}} \text{gen}_{p,q} \right] \quad (4.11)$$

by optimising over $[\text{cap}^{\text{gen}}, \text{gen}_{p,q} \mid p=1\dots N_{\mathcal{P}}, q=1\dots N_{\mathcal{Q}}]$, subject to

$$\text{gen}_{p,q} = \hat{d}_{\mathcal{A}(p),q} \quad \text{for all } p = 1\dots N_{\mathcal{P}}, q = 1\dots N_{\mathcal{Q}} \quad (4.12)$$

$$\text{gen}_{pq} \in [0, \text{cap}^g] \quad \text{for all } p = 1\dots N_{\mathcal{P}}, q = 1\dots N_{\mathcal{Q}}. \quad (4.13)$$

This is symmetric (in the constraints and objective function) in any two generation levels mapped to the same representative period — i.e. any $\text{gen}_{p_1,q}$ and $\text{gen}_{p_2,q}$ for which $\mathcal{A}(p_1) = \mathcal{A}(p_2)$. Their values are hence equal and we may (without any loss of generality) use the mapping

$$\text{gen}_{p,q} \mapsto \text{gen}_{\mathcal{A}(p),q} \quad \text{for all } p = 1\dots N_{\mathcal{P}}, q = 1\dots N_{\mathcal{Q}}. \quad (4.14)$$

These reduce the number of decision variables from $N_{\mathcal{P}}N_{\mathcal{Q}}$ to $N_{\mathcal{P}_r}N_{\mathcal{Q}}$ — one per *representative* period. They are weighted in the objective function by their number of occurrences in the original time series. They also reduce the number of constraints for (4.12) and (4.13) to $N_{\mathcal{P}_r}N_{\mathcal{Q}}$ each (since duplicate constraints on the same representative period can be removed).

In this way, time series input aggregation reduces the number of decision variables, constraints and hence computational cost. The same logic applies to models with more technologies regions or constraints, which all appear in proportion to the number of representative periods.

4.1.3 Literature review

4.1.3.1 Time series aggregation

The power system modelling community employs various time series aggregation schemes, as summarised by Hoffmann et al. [2020] and Teichgraeber and Brandt [2022]. Merrick [2016] discusses performance criteria, noting that load duration curves (Section 3.4.2.1) and correlations between time series should be preserved. Period lengths may differ; single time step blocks can represent a large diversity of scenarios in a relatively short time series, while e.g. daily periods preserve short-term dependence structures such as intra-day storage or ramping (constraints on rates of change of generation levels).

Representative periods are determined in various ways. Welsch et al. [2012] create one week- and weekend-day for each season. De Sisternes and Webster [2013] and Poncelet et al. [2017] select periods by minimising the distortion of aggregated

duration curves (see Section 3.4.2.1), exhaustively checking each combination. Many studies cluster vectors of daily time series values; Nahmmacher et al. [2016], Poncelet et al. [2017], Härtel et al. [2017], Pfenninger [2017] and Kotzur et al. [2018a] evaluate variations on this approach.

4.1.3.2 *A posteriori* methods

Most time series aggregation schemes are what Hoffmann et al. [2020] call *a priori*: they use information about the input time series but not the power system model. Such schemes generate the same aggregation for any model, regardless of e.g. whether generation is 0% or 100% renewable. This can be suboptimal, since reducing time series approximation errors (e.g. the deviation of duration curves) need not lead to more accurate estimates of optimal design or cost [Gonzato et al., 2021, Teichgraeber and Brandt, 2022, Wogrin, 2022].

In contrast, *a posteriori* (also known as *adaptive*) methods tailor subsamples to the power system. For example, Fitiwi et al. [2015] view the grid’s power flow as an image and cluster its moments. Sun et al. [2019] cluster the vectors of investment costs (determined by running the planning model on each individual day) instead of the time series data itself. Zhang et al. [2022] do the same with other model outputs, such as the optimal installed capacities.

Another class of approaches uses the model’s operational variables. First, an optimal design estimate is obtained from a planning model with *a priori* aggregation. An operation problem with this design, sometimes called the *production cost model*, is subsequently run across the full time series. The resultant operational variables — generation, transmission and storage levels — are subsequently used to improve aggregation. For example, Bahl et al. [2018] and Teichgraeber et al. [2021] alternate between the planning problem and operation problem, iteratively including days with unmet demand until unserved energy falls below a threshold or an optimality bound is reached.

4.1.4 Conclusions and implications

The studies above arrive at some common conclusions. The first is that, as illustrated in Figure 4.2, constructing representative periods by averaging (e.g. using, for each hour in a day, its average value across a season) performs poorly since it reduces extremes and variability [Bothwell and Hobbs, 2018]. Furthermore, no particular algorithm dominates others, as relative performance often depends on

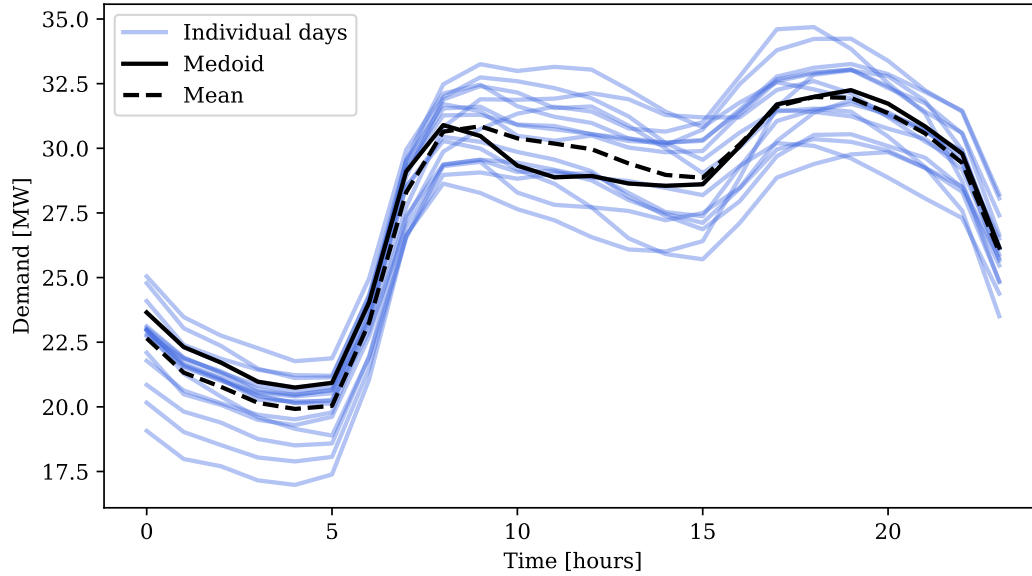


Figure 4.2: Reduction in extremes and variability from *mean*- and *medoid*-based methods. The medoid reduces variability to a smaller degree.

the underlying PSM [Kotzur et al., 2018a](#).

Preserving relevant extremes events, such as those with high demand but low renewable output, is important for accurate design estimates. Removing such events typically leads to (1) an underestimation of required backup (dispatchable) generation capacity and (2) an over-estimation of the reliability of variable renewables. This, in turn, causes an under-estimation of system costs and system designs unable to meet demand across the full range of demand and weather scenarios [Härtel et al., 2017](#), [Hoffmann et al., 2020](#). This is a result of cost minimisation — cost-optimal design across a time series with peak demand 50MW cannot meet a 60MW demand event — and affects both averaging- and clustering-based methods, which truncate extremes by replacing individual periods by averages (Figure [4.2](#)).

Users sometimes make heuristic adjustments to preserve extreme events, such as including the day with peak demand or minimum renewable potential [Pfenninger, 2017](#), [Hoffmann et al., 2020](#). Furthermore, *medoid*-based methods, in which the representative period is the *real* period closest to the cluster mean, avoids some reduction in variability, as shown in Figure [4.2](#).

4.2 Contribution

This chapter’s contribution is as follows. Sections [2.4](#)[2.5](#) indicate that long demand and weather samples should be considered to avoid incorrect quantitative and qualitative conclusions from power system planning models, but that solving the associated optimisation problem is often computationally unfeasible. Furthermore, established time series aggregation approaches can induce significant errors due to their failure to preserve relevant extremes (Section [4.1.4](#)). We use *extreme* to refer to time steps in which meeting demand is difficult, e.g. with a high demand but low renewable output.

We introduce a method, called *importance subsampling*, that systematically identifies relevant extremes and includes them in subsamples. As a result, it estimates long-term optimal system design at reduced computational cost and with lower levels of error than previous aggregation methods. It allows the consideration of much longer (e.g. multi-decadal) demand and weather data in power system planning models — as is essential to analysis of future climate or extreme events which occur infrequently — with much lower levels of error than currently available methods. A simulation study demonstrates the method’s improvements.

Since the publication of our papers [\[Hilbers et al., 2019, 2020\]](#), a number of similar methods have been proposed, either independently or with reference to our work. These include some *a posteriori* methods discussed in Section [4.1.3.2](#) which we include but were not available at the time of our publications.

4.3 Methods

4.3.1 Overview and intuition

Importance subsampling tailors aggregation to the underlying power system model by explicitly including a set of system-relevant extremes. This avoids some pitfalls discussed in Section [4.1.4](#), namely the under-estimation of required backup (dispatchable) generation capacity and system costs.

The key innovation is how system-relevant extremes are identified. This involves a trade-off between different time series. For example, in a model with unknown wind generation capacity, it is unclear whether a 50 MW demand, 0.1 wind generation potential event requires more dispatchable capacity than one with 60 MW demand and 0.5 wind generation potential. This problem becomes more difficult in multi-

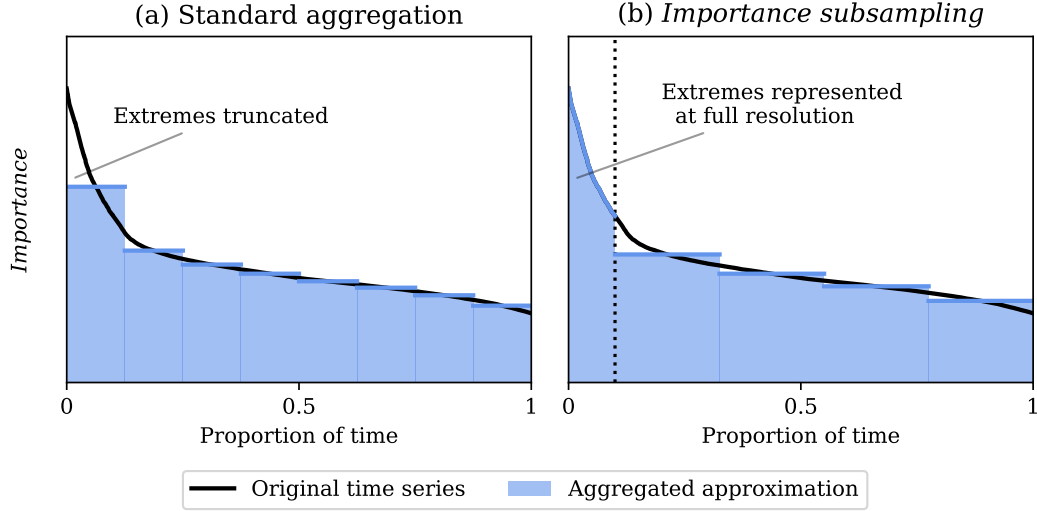


Figure 4.3: Impacts of time series aggregation. The black line is the *duration curve* (Section 3.4.2.1) of the *importance* across the original (full) time series. (a) Standard aggregation truncates extremes with a high *importance*. (b) *Importance subsampling* represents extremes at full resolution.

region models, since the peak demand in one region may not occur at the same time in another. We identify system-relevant extremes by their *importance*, a user-defined proxy. In many applications, a time step's *generation cost* is a good candidate; we discuss this and others in Section 4.3.3.

Calculating a time period's *importance* requires a model's operational variables and hence, indirectly, the system design. For example, the generation cost (an operational variable) to meet a 60 MW, 0.5 wind generation potential event is zero with 120 MW of wind capacity but requires 60 MW of e.g. gas (with a nonzero generation cost) if no wind capacity is present. Ideally, we use the actual optimal system design — as determined across the full time series without aggregation — but this is the output being estimated and hence unknown. This is a cyclical dependency: we want a time period's *importance* to improve time series aggregation for the optimal design estimate, but we need the design to determine the *importance*.

For this reason, we propose a two-stage approach. We run a preliminary planning model with a standard aggregation scheme for a rough estimate of optimal system design. We use this to determine the *importance* of each time period, and include those with the highest value (a proxy for the most *extreme* periods) in an aggregated time series. Finally, we run the planning model on this aggregated sample for a second, hopefully more accurate, estimate of optimal system design.

To determine *importance* values, we run an operational problem (Section 2.3.1)

across the full time series. In these problems, we provide a design $\hat{\mathbf{D}}$ and determine the cost-optimal system operation $(\hat{\mathbf{O}}_p)_{p \in \mathcal{P}}$ (generation and transmission levels in each period p):

$$(\hat{\mathbf{O}}_p)_{p \in \mathcal{P}} = \text{PSM}_{\text{operate}}((\boldsymbol{\xi}_p)_{p \in \mathcal{P}} \mid \hat{\mathbf{D}}). \quad (4.15)$$

4.3.2 Algorithm: *importance subsampling*

Inputs:

- $(\boldsymbol{\xi}_p)_{p \in \mathcal{P}}$: demand and weather time series, length $n_{\mathcal{P}}$ periods
- $n_{\mathcal{A}}$: number of unique representative periods to aggregate into
- $p_e \in [0, 1]$: proportion of *extreme* representative days in sample
- IMP: real-valued *importance* function, from operational variables $\mathbf{O}_p$

Steps:

1. Get preliminary optimal design estimate $\hat{\mathbf{D}}_{\mathcal{A}_0}$:

- (a) Aggregate $(\boldsymbol{\xi}_p)_{p \in \mathcal{P}}$ into $(\hat{\boldsymbol{\xi}}_{\mathcal{A}_0(p)})_{p \in \mathcal{P}}$, with $n_{\mathcal{A}}$ unique representative periods, via an *a priori* scheme.
- (b) Solve planning problem:

$$\hat{\mathbf{D}}_{\mathcal{A}_0} = \text{PSM}_{\text{plan}}((\hat{\boldsymbol{\xi}}_{\mathcal{A}_0(p)})_{p \in \mathcal{P}}). \quad (4.16)$$

2. Create *importance subsample*:

- (a) Determine system operation across full time series:

$$(\hat{\mathbf{O}}_p)_{p \in \mathcal{P}} = \text{PSM}_{\text{operate}}((\boldsymbol{\xi}_p)_{p \in \mathcal{P}} \mid \hat{\mathbf{D}}_{\mathcal{A}_0}). \quad (4.17)$$

- (b) Calculate *importance* of each period:

$$\text{imp}_p = \text{IMP}(\hat{\mathbf{O}}_p) \quad \text{for all } p \in \mathcal{P}. \quad (4.18)$$

- (c) Create *importance subsample* $(\hat{\boldsymbol{\xi}}_{\mathcal{A}_1(p)})_{p \in \mathcal{P}}$, $n_{\mathcal{A}}$ periods, with:

- $p_e n_{\mathcal{A}}$ *extreme* periods, with highest *importance* in $\mathcal{P}$
- $(1 - p_e)n_{\mathcal{A}}$ *regular* periods (via an *a priori* scheme) aggregated from the $n_{\mathcal{P}} - p_e n_{\mathcal{A}}$ remaining.

Weight periods according to their relative proportions in $\mathcal{P}$.

3. Get final optimal design estimate $\hat{\mathbf{D}}_{\mathcal{A}_1}$:

$$\hat{\mathbf{D}}_{\mathcal{A}_1} = \text{PSM}_{\text{plan}}((\hat{\boldsymbol{\xi}}_{\mathcal{A}_1(p)})_{p \in \mathcal{P}}). \quad (4.19)$$

Output: $\hat{\mathbf{D}}_{\mathcal{A}_1}$: optimal design estimate

4.3.3 Remarks

Importance subsampling is an *a posteriori* (Section 4.1.3.2) scheme that customises the aggregation to the model. An *a priori* scheme would use $\hat{\mathbf{D}}_{\mathcal{A}_0}$ in Step 1 as the final estimate. We instead use it to create an additional sample, customised to the underlying planning model, for a second design estimate $\hat{\mathbf{D}}_{\mathcal{A}_1}$.

The one-dimensional *importance* function IMP used in step 2(b) identifies *extreme* periods to include in the aggregated time series. In our case study (Section 4.4), we use a period's total *generation cost*. This identifies *extreme* periods that require peak levels of generation, since expensive measures such as peaking plants are only used in settings where there are otherwise supply shortages. Other candidates include electricity price or unmet demand levels. Knowledge of the planning model can motivate a choice of *importance* function.

The hyperparameter p_e , the proportion of the $n_{\mathcal{A}}$ -period subsample reserved for *extreme* periods, represents a trade-off: increasing p_e means a larger selection of *extreme* events and a smaller number of *regular* ones. We find $p_e = \frac{1}{3}$ to work well. With this value, a sample with 120 representative periods includes the 40 periods in the full time series with the highest *importance*, which are included without any subsampling (see Figure 4.3).

Estimating optimal system design using *importance subsampling* requires three model runs (with correspondence to Section 4.3.2 shown in parentheses):

1. Planning model on $n_{\mathcal{A}}$ representative periods $(\hat{\xi}_{\mathcal{A}_0(p)})_{p \in \mathcal{P}}$ (Step 1).
2. Operation model on full time series $(\xi_p)_{p \in \mathcal{P}}$, $n_{\mathcal{P}}$ periods (Step 2).
3. Planning model on $n_{\mathcal{A}}$ representative periods $(\hat{\xi}_{\mathcal{A}_1(p)})_{p \in \mathcal{P}}$ (Step 3).

The total computing time is roughly twice that for an *a priori* method with the same sample length $n_{\mathcal{A}}$. This is because the two planning problems (both length $n_{\mathcal{A}}$) dominate solution times. Despite its longer simulation length ($n_{\mathcal{P}}$ vs. $n_{\mathcal{A}}$ periods), the operation problem runs faster since the design is fixed. Memory requirements are the same as for standard methods; it can be reallocated between simulations, and operation problem memory requirements are independent of sample length due to decomposition into horizons and windows (Section 3.4.1.3).

The method relies on a number of implementation details, which we discuss here; the reader can also look at the code directly at doi:10.5281/zenodo.3841154. Firstly, the aggregation in Steps 1(a) and 2(c) can use any *a priori* scheme. In our case study, we use *k-medoids* clustering, implemented as follows:

1. Normalise each time series to $[0, 1]$ by a linear transform (map minimum to zero, maximum to one). Create one vector per period as in (4.2).
2. Cluster these vectors via Lloyd’s k -means algorithm [Lloyd, 1982].
3. Replace each period by its *medoid*: the period closest (in normalised space with the Euclidean metric) to the cluster mean.
4. Undo the normalisation to return dimensional time series values.

Periods in $(\hat{\xi}_{\mathcal{A}_1(p)})_{p \in \mathcal{P}}$ are weighted in the objective function to account for their proportions in the full time series. For example, the 30 most extreme days in a 30-year sample represent one day a year in the full time series but 30 days a year in a one-year subsample. We weight them to correct these biases. Finally, the period length must also be chosen. Single time step (e.g. hourly) blocks represent a larger diversity of scenarios with the same number of time steps, while longer periods preserve chronology inside blocks.

The generation cost is the value of the second stage of the planning problem (see Section 2.7). In this way, we may view this method as importance sampling that prioritises events with a high contribution to the second stage cost. Through this oversampling of rare but high-impact events, we reduce the variance of our estimators as described in Appendix A.2. Similar ideas have been applied in two-stage problems by [Dantzig and Glynn, 1990], [Infanger, 1992] and, more recently, [Parpas et al., 2015]. We call our method *importance subsampling* instead of *importance sampling* since we sample an existing time series instead of a probability distribution.

4.4 Simulation studies

4.4.1 Overview

In this section, we evaluate the method’s performance on the *CLP* and *MILP* models introduced in Chapter 3. We do this as follows. First, we obtain an optimal design estimate $\hat{\mathbf{D}}_{\mathcal{A}}$ using a planning model with either *importance subsampling* or *k-medoids* clustering, representing the current industry standard *a priori* method. We also calculate the optimal design across the full time series without any aggregation, which we view as the ground truth to be approximated. We compare estimated optimal designs (vectors of installed capacities) with the ground truth optima, and seek aggregation that makes them as close together as possible.

We furthermore use unmet demand as an evaluation metric. In our planning models, we enforce that all demand is met by generation. For this reason, design estimates lead to zero unmet demand across the sample they are optimised across, and the true optimum design (calculated across the full time series $(\xi_p)_{p \in \mathcal{P}}$ without aggregation) has no unmet demand. However, a design estimate $\hat{\mathbf{D}}_{\mathcal{A}}$ (calculated from a planning model *with* time series aggregation) may lead to some unmet demand across $(\xi_p)_{p \in \mathcal{P}}$ since some events are removed under aggregation and hence do not appear in the aggregated time series. We calculate the magnitude of this unmet demand by running an operation problem with design $\hat{\mathbf{D}}_{\mathcal{A}}$ (while allowing unmet demand to ensure solution feasibility) across the full time series $(\xi_p)_{p \in \mathcal{P}}$. In this case, we seek aggregation for which the design estimate $(\hat{\xi}_{\mathcal{A}(p)})_{p \in \mathcal{P}}$ has as little unmet demand across $(\xi_p)_{p \in \mathcal{P}}$ as possible.

Running the planning model without allowing unmet demand makes optimal design estimates sensitive to sample maxima, since installed capacities must meet the peak demand event. We could also allow unmet demand in the planning stage; this would reduce the dependence on sample maxima and relate more closely to probabilistic security of supply standards (e.g. <0.1% of energy unserved). Such an approach also allows a trade-off between unmet demand and generation capacity which is rarely used. However, it removes the interpretation of unmet demand as the degree of infeasibility of a design estimate across the full time series. It also means target unmet demand is no longer zero, but a positive value. For these reasons, in order to use unmet demand directly as an error metric (as opposed to e.g. *excess unmet demand* — the value above what is cost-optimal), we present results *without* allowing unmet demand in planning models; the reader should note that optimal designs may be sensitive to sample maxima (e.g. peak demand) in the time series used in planning problems.

4.4.2 Setup

Details for the simulations are provided in Table [4.1](#). We use daily periods for all simulations. For the *CLP* model, optimal design across 2008-2017 serves as the target for simulations with $n_{\mathcal{A}}=48$ and 90 representative days. For the *MILP* model, the integer constraints impose (much) longer solution times and the optimal design across 2017 is estimated using $n_{\mathcal{A}}=24$ and 48 representative days. Since both *k*-medoids clustering and *importance subsampling* contain randomness from the initialisation of cluster means, we determine the distribution of model outputs

Scheme	Time series aggregation		Computing	
	$n_{\mathcal{A}}$ [days]	$p_e n_{\mathcal{A}}$ [days]	Time [min]	Memory [MB]
<i>CLP</i> model, full time series 2008-2017				
None (target)	3650	-	340	15121
k -medoids clustering	48	-	0.5	393
Importance subsampling	48	16	0.5+11+0.5	742
k -medoids clustering	90	-	1	567
Importance subsampling	90	30	1+11+1	753
<i>MILP</i> model, full time series 2017				
None (target)	365	-	2926	1598
k -medoids clustering	24	-	2	270
Importance subsampling	24	8	2+2+2	823
k -medoids clustering	48	-	9	362
Importance subsampling	48	16	9+2+9	841

Table 4.1: Simulations in Section 4.4. $n_{\mathcal{A}}$: sample length, $p_e n_{\mathcal{A}}$: number of *extreme* days in sample. *Importance subsampling* solution times are disaggregated into first planning, operational run and second planning runs.

across 100 subsamples.

The implementation of *importance subsampling* (Section 4.3.2) is as follows. k -medoids clustering is used in steps 1(a) and 2(c). Each time series is normalised to lie between 0 and 1. For the *importance* function in step 2(b), we use a period's total (sum) *generation cost*. We use $p_e = \frac{1}{3}$; each sample contains twice as many *regular* as *extreme* days.

4.4.3 Results

Figure 4.4 shows the distribution of estimated optimal capacities, peak capacity shortage and unmet demand across the test simulations. For optimal capacity estimates, *importance subsampling* generally has superior performance to k -medoids clustering. For example, the latter usually underestimates (overestimates) optimal peaking (wind) capacity. These errors align with previous studies (Section 4.1.4): an underestimation of optimal dispatchable capacity and an overestimation of optimal variable renewable capacity. Such biases are not present for *importance subsampling*. Furthermore, the variability across samples is larger with k -medoids clustering than *importance subsampling*. Statistically, the estimates of optimal system design have both a lower bias and variance than those obtained using the *a priori* clustering method. For both k -medoids clustering and *importance subsampling*, design esti-

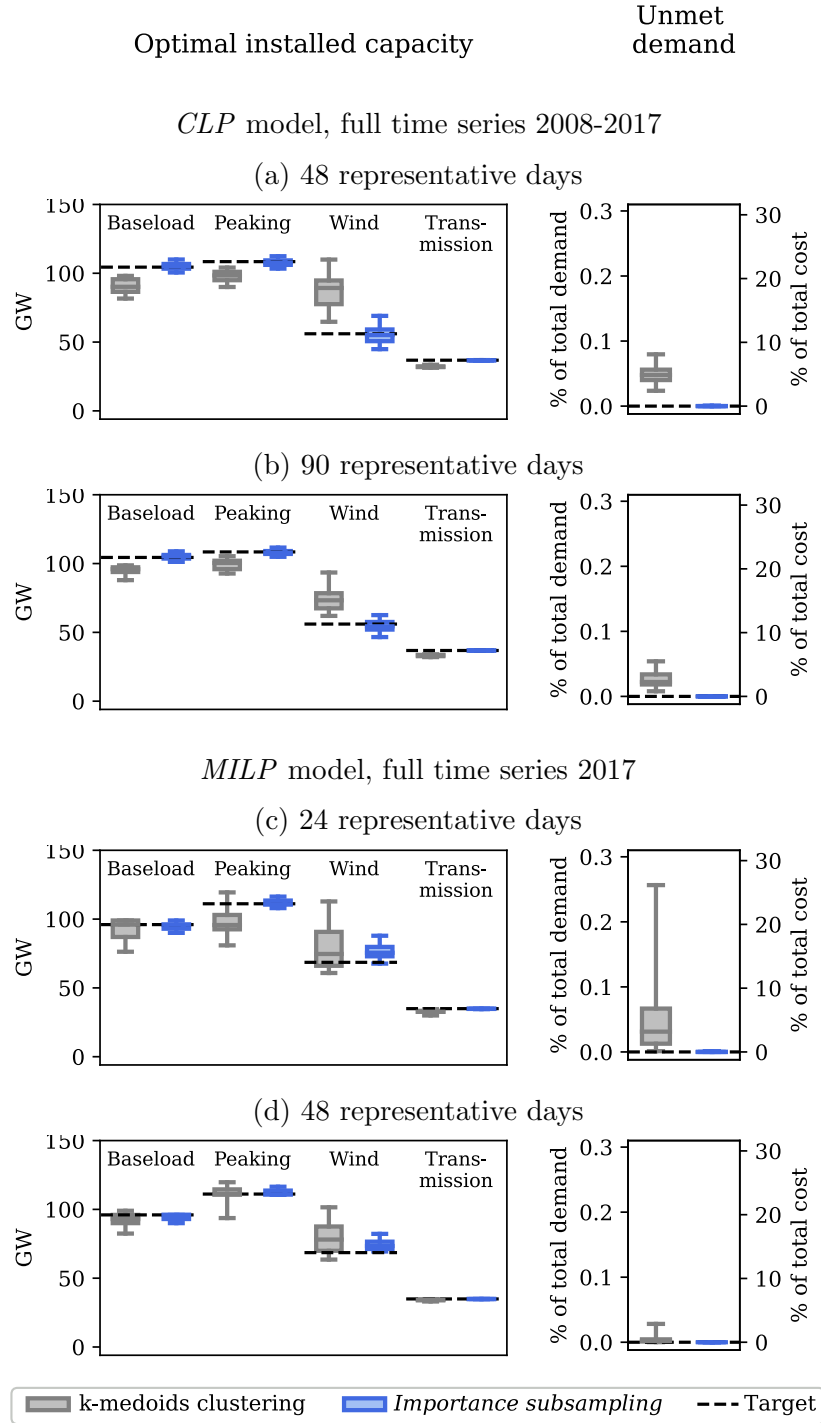


Figure 4.4: Distribution of optimal capacities and unmet demand across 100 samples generated by *k-medoids clustering* and *importance subsampling*. Box-and-whiskers: 2.5%, 25%, median, 75% and 97.5% percentiles. Dashed black line: optima across the full time series (target).

mates become more accurate as the number of representative days n_A is increased.

As a result, levels of unmet demand are lower (almost zero) for *importance subsampling* than for *k-medoids*. A system design based on 48 representative days *k-medoids* clustered from 10 years (Figure 4.4(a)) fails to meet 0.05% of demand on average, compared to virtually none for a system designed using *importance subsampling*. This has financial implications; assigning a value of lost load of £6000/MWh [Elexon, 2018], implies an increase in system cost of around 1% per each 0.01% of demand that is unmet.

4.4.4 Discussion

The performance gain of *importance subsampling* comes preserving extreme events. By design, *k-medoids* clustering is equivalent to using $\hat{\mathbf{D}}_{A_0}$ in Step 1 (Section 4.3.2) as the optimal design estimate. Hence, improvements come from Steps 2-3, in which extreme periods are identified and included. It is to be expected that removing these events results in the underestimation (overestimation) of peaking (wind) capacity observed in Figure 4.4: samples not including high demand, low wind events underestimate the amount of required backup generation capacity while overestimating the reliability of wind power.

The performance gap appears more strongly when aggregating from longer time series, as seen by comparing Figure 4.4(a) and (d). This is because longer, more diverse samples are more difficult to represent with the same number of representative days. Hence, this performance gap is expected to grow as longer time series (e.g. multiple decades) are sampled from.

Importance subsampling can estimate optimal design across much longer simulations than those used here. For example, we could estimate a 25-year output using a one-year subsample. The use of shorter samples is only to make validation of the method possible by comparing with the ground truth determined by a brute force model run on the full time series; a direct 25-year simulation is computationally unfeasible.

The use of short samples in this manner distorts the computational expense. In realistic planning models, computational expense for *importance subsampling* is dominated by the two planning runs. This is because solution times for the operational model scale linearly in the number of time steps while that for a planning run scales faster. For example, for the MILP model, a one-year estimate of 25-year optimal design requires around 48 hours for each planning run but just 48 minutes

for the operational run across the full 25 years.

In our case study, we do not allow unmet demand in the planning models; the resultant system designs meet 100% of demand in the sample they are calculated across. This gives unmet demand (in the operation problems used as evaluation) an interpretation as the degree of unfeasibility of the estimated system design. We can also allow unmet demand in the planning stage, in which case an economic tradeoff is made between unserved energy and capacity that is rarely used. In this case, estimates from load duration curves (Section 3.4.2.1, a proxy calculated from the relative costs of generation technologies and the value of lost load) indicate leaving roughly 0.007% of demand to be unmet is cost-optimal. The resultant system designs would have lower peaking capacities (since this is the technology that matches peak demand). *Importance subsampling* would still perform better than k -medoids clustering, for which levels of unmet demand greatly exceed the cost-optimal 0.007% (similar to those observed in Figure 4.4).

4.5 Conclusions

This chapter’s contribution is the *importance subsampling* method. It can reliably estimate long-term planning model outputs (considering multiple decades of demand and weather data) at greatly reduced computational cost, outperforming industry standard *a priori* methods. As discussed in Section 2.5, the consideration of such long samples is essential for robust power system planning under natural climate/weather variability. It also opens the door to the consideration of climate model data — which typically spans multiple decades — in power system planning models (see e.g. [Wohland et al., 2017, Craig et al., 2019, van Zuijlen et al., 2019]).

Importance subsampling is an *a posteriori* method (see Section 4.1.3) in which properties of the model are used to improve aggregation. The key innovation is a systematic identification and inclusion of relevant *extreme* events in subsamples through a user-defined *importance* function. This function may be tailored to the specific application; the generation cost and electricity price are candidates. Subsequent studies with similar methods, such as those by [Teichgraeber et al., 2021], have used unmet demand. The systematic identification of relevant extremes may also be used in its own right, for example by analysing the identified *extreme* days at higher spatiotemporal resolution.

Importance subsampling uses two planning runs to obtain $\hat{\mathbf{D}}_{\mathcal{A}_0}$ and $\hat{\mathbf{D}}_{\mathcal{A}_1}$. We can, however, iterate steps 2 and 3, iteratively creating more design estimate $\hat{\mathbf{D}}_{\mathcal{A}_i}$

for $i > 1$. This approach, where we alternate between planning and operational model runs, is similar to those introduced by [Bahl et al. \[2018\]](#) and [Teichgraeber et al. \[2021\]](#), which iteratively include days with unmet demand.

The key extension for this method is adjusting it to work with temporal dependencies such as inter-day storage. Like many other temporal aggregation approaches (Section [4.1.3](#)), *importance subsampling* reorders periods, breaking their chronology. This makes it unsuitable for models with inter-day storage. We extend the method to work with storage in the next chapter.

5. Efficient design estimates with storage

In this chapter, we generalise *importance subsampling* (introduced in the previous chapter) to planning models with inter-day storage, which necessitates the preservation of chronology. The goal is unchanged: estimate cost-optimal system design across a long time series at reduced computational cost. The algorithm has been published in the following paper:

- [\[Hilbers et al., 2022\]](#) AP Hilbers, DJ Brayshaw, A Gandy (2022). Reducing climate risk in energy system planning: a posteriori time series aggregation for models with storage. *Applied Energy*.

Models, data, methods and code are available at doi:[10.5281/zenodo.7178301](https://doi.org/10.5281/zenodo.7178301).

This chapter is structured as follows. Section [5.1](#) provides background on the problem, the implications of intertemporal dependencies, and previous approaches. Section [5.2](#) describes our contribution. Section [5.3](#) introduces the method and Section [5.4](#) provides a case study of its application on the test PSMs of Chapter [3](#). Section [5.5](#) makes concluding remarks.

5.1 Background

5.1.1 Context: time series aggregation

Our research context mirrors Chapter [4](#). Determining cost-optimal system design $\hat{\mathbf{D}}_{\mathcal{T}}$ across the longest available demand and weather time series $(\xi_t)_{t \in \mathcal{T}}$ is often computationally unfeasible. We estimate it by $\hat{\mathbf{D}}_{\mathcal{A}}$, calculated across a shorter sample created by *time series aggregation*.

The aggregation methods discussed in Chapter [4](#) are unsuitable for models with inter-period storage since they distort the chronology on which storage patterns depend. In this section, we explain why this is and discuss possible solutions. In particular, we show that time series aggregation for models with storage require both (1) the preservation of chronology information and (2) aggregation of both time series inputs *and* storage decision variables.

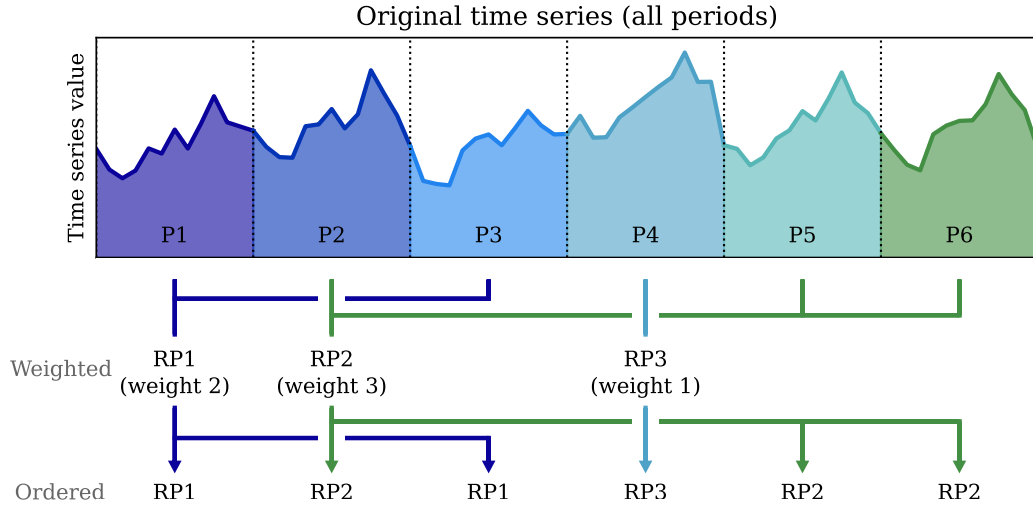


Figure 5.1: Time series aggregation from six periods (P1-6) to three *representative periods* (RP1-3), either *weighted* (appearing once, weighted by number of occurrences in full time series) or *ordered* (in same order as full time series).

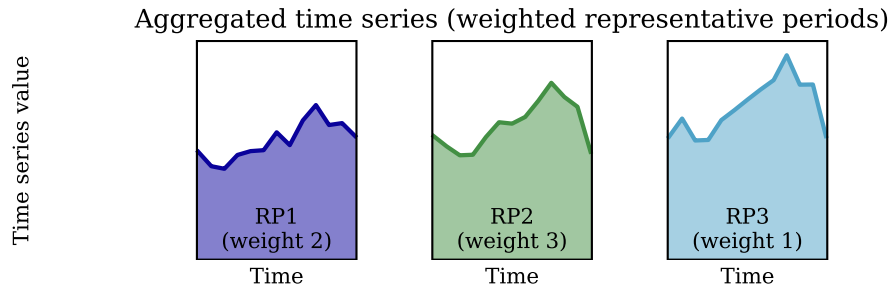


Figure 5.2: Time series after aggregation into *weighted* representative periods.

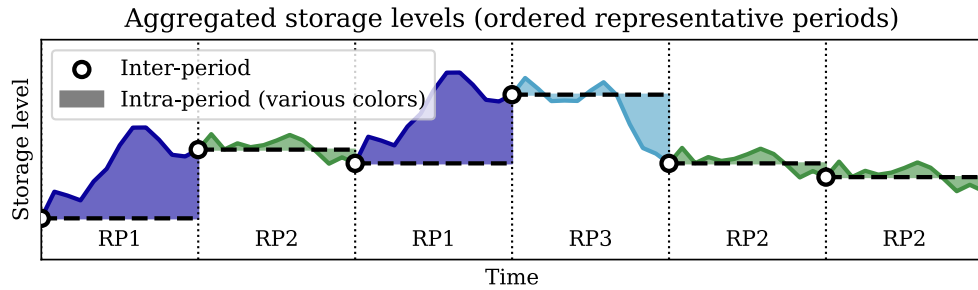


Figure 5.3: Storage levels after aggregation into *ordered* representative periods and decomposition into *inter-period* (level at start of period) and *intra-period* (change compared to start of period) levels. Intra-period contributions are equal across replications of the same period but inter-period levels may differ.

5.1.2 Mathematical background

5.1.2.1 Time series aggregation

We reuse notation from Chapter 4, with $(\xi_p)_{p \in \mathcal{P}}$ a time series partitioned into periods. Recall that *time series aggregation* $\mathcal{A} : \mathcal{P} \rightarrow \mathcal{P}$ maps $(\xi_p)_{p \in \mathcal{P}}$ to *representative periods* $(\hat{\xi}_{\mathcal{A}(p)})_{p \in \mathcal{P}}$, where $N_{\mathcal{P}_r}$, the number of unique periods in $\{\mathcal{A}(p) \mid p \in \mathcal{P}\}$, is smaller than $N_{\mathcal{P}}$, the number of periods in $\{p \mid p \in \mathcal{P}\}$.

We distinguish two classes of representative periods. The first is *weighted*, where each period appears once, weighted by its number of occurrences; in this case their chronology is lost. The second is *ordered*, where we replace each period by its representative counterpart but keep them in order. For the example from six to three representative periods from Section 4.1.2, these are:

$$\begin{aligned}
 (\xi_p)_{p \in \mathcal{P}} &= (\xi_1 \quad \xi_2 \quad \xi_3 \quad \xi_4 \quad \xi_5 \quad \xi_6) \\
 &\quad \downarrow \\
 \text{ordered:} &\quad (\hat{\xi}_1 \quad \hat{\xi}_2 \quad \hat{\xi}_1 \quad \hat{\xi}_3 \quad \hat{\xi}_2 \quad \hat{\xi}_2) \\
 \text{weighted:} &\quad (\hat{\xi}_1 \text{ (weight 2)} \quad \hat{\xi}_2 \text{ (weight 3)} \quad \hat{\xi}_3 \text{ (weight 1)}).
 \end{aligned} \tag{5.1}$$

These are illustrated in Figures 5.1, 5.3 and discussed in the next section.

5.1.2.2 Impact on planning problem

In this section, we examine the impact of time series aggregation on power system planning models with storage. We show that, to reduce the number of decision variables and constraints (and hence computational cost), we must aggregate both input time series and storage (dis)charging decisions.

We mirror the case without storage (Section 4.1.2) and use a similar toy model. Consider a time series split into $N_{\mathcal{P}}$ periods with $N_{\mathcal{Q}}$ time steps each, and let (p, q) index time step q of period p . The model has one generation technology with install cost C^{gen} (£/MW per time step of lifetime), generation cost F^{gen} (£/MWh), installed capacity cap^{gen} (MW) and generation levels $g_{p,q}$ (MWh). It also has storage with install cost C^{sto} (£/MWh per time step of lifetime), energy storage capacity cap^{sto} (MWh), storage level $\text{sto}_{p,q}$ and storage charging $\text{ch}_{p,q}$ (MWh). The demand $d_{p,q}$ is the only time series input, so that $\xi_p = [d_{p,1}, \dots, d_{p,N_{\mathcal{Q}}}]$ contains inputs per

period. The planning problem is:

$$\min \left[\underbrace{N_{\mathcal{P}}N_{\mathcal{Q}}(C^{\text{gen}}\text{cap}^{\text{gen}} + C^{\text{sto}}\text{cap}^{\text{sto}})}_{\text{install cost}} + \underbrace{F^g \sum_{p=1}^{N_{\mathcal{P}}} \sum_{q=1}^{N_{\mathcal{Q}}} \text{gen}_{p,q}}_{\text{generation cost}} \right] \quad (5.2)$$

by optimising over $[\text{cap}^{\text{gen}}, \text{cap}^{\text{sto}}, \text{gen}_{p,q} \mid p=1\dots N_{\mathcal{P}}, q=1\dots N_{\mathcal{Q}}]$, subject to

$$\text{gen}_{p,q} = d_{p,q} + \text{ch}_{p,q} \quad \text{for all } p = 1\dots N_{\mathcal{P}}, q = 1\dots N_{\mathcal{Q}} \quad (5.3)$$

$$\text{sto}_{1,0} = 0 \quad (5.4)$$

$$\text{sto}_{p,q+1} = \text{sto}_{p,q} + \text{ch}_{p,q+1} \quad \text{for all } p = 1\dots N_{\mathcal{P}}, q = 0\dots N_{\mathcal{Q}}-1 \quad (5.5)$$

$$\text{sto}_{p+1,0} = \text{sto}_{p,N_{\mathcal{Q}}} \quad \text{for all } p = 1\dots N_{\mathcal{P}}-1 \quad (5.6)$$

$$\text{gen}_{pq} \in [0, \text{cap}^{\text{gen}}] \quad \text{for all } p = 1\dots N_{\mathcal{P}}, q = 1\dots N_{\mathcal{Q}} \quad (5.7)$$

$$\text{sto}_{pq} \in [0, \text{cap}^{\text{sto}}] \quad \text{for all } p = 1\dots N_{\mathcal{P}}, q = 1\dots N_{\mathcal{Q}} \quad (5.8)$$

where $\text{sto}_{p,0}$ is the storage level at the start of period p and $\text{sto}_{p,q}$ is its level at the end of the q th time step. The constraints specify that generation equals demand plus storage charging (5.3), empty initial storage (5.4), lossless storage level updates within periods (5.5), linking of storage levels between periods (5.6), and generation and storage bounds (5.7)-(5.8). Storage levels are not independent decision variables (they are rolling sums of $\text{ch}_{p,q}$), but make the constraints easier to interpret. All constraints apply to single periods except (5.6), which *link* storage levels between periods; the storage level at the start of a period is the same as at the end of the previous one.

For time series aggregation, we implement the method introduced by Kotzur et al. (2018b) and illustrated in Figure 5.3. First, we decompose storage levels into *inter*- and *intra*-period contributions via

$$\text{sto}_{p,q} = \text{sto}_p^{\text{inter}} + \text{sto}_{p,q}^{\text{intra}} \quad \text{for all } p = 1\dots N_{\mathcal{P}}, q = 0\dots N_{\mathcal{Q}} \quad (5.9)$$

$$\text{sto}_{p,0}^{\text{intra}} = 0 \quad \text{for all } p = 1\dots N_{\mathcal{P}} \quad (5.10)$$

where the second stipulates that inter-period levels are those at the start of each period. This decomposition does not change the problem (we have made no new assumptions, and the solution is the same) but it decomposes decision variables and constraints into inter- vs. intra-period scales. Second, we make aggregation assumptions on both the input time series (demand levels) and storage (dis)charge

decisions:

$$d_{p,q} \mapsto \hat{d}_{\mathcal{A}(p),q} \quad \text{for all } p = 1 \dots N_{\mathcal{P}}, \quad q = 1 \dots N_{\mathcal{Q}} \quad (5.11)$$

$$\text{ch}_{p,q} \mapsto \hat{\text{ch}}_{\mathcal{A}(p),q} \quad \text{for all } p = 1 \dots N_{\mathcal{P}}, \quad q = 1 \dots N_{\mathcal{Q}}. \quad (5.12)$$

The aggregation of time series inputs (demand levels) is equivalent to the case without storage. However, we now also aggregate storage (dis)charge decisions by replacing all instances in the same representative period by a single decision variable. These can still be chosen to minimise costs (unlike the demand levels, which are fixed), but are re-used across replications of the same representative period. The constraints become:

$$\text{gen}_{p,q} = \hat{d}_{\mathcal{A}(p),q} + \hat{\text{ch}}_{\mathcal{A}(p),q} \quad \text{for all } p = 1 \dots N_{\mathcal{P}}, \quad q = 1 \dots N_{\mathcal{Q}} \quad (5.13)$$

$$\text{sto}_1^{\text{inter}} = 0 \quad (5.14)$$

$$\text{sto}_{p+1}^{\text{inter}} = \text{sto}_p^{\text{inter}} + \text{sto}_{p,N_{\mathcal{Q}}}^{\text{intra}} \quad \text{for all } p = 1 \dots N_{\mathcal{P}} \quad (5.15)$$

$$\text{sto}_{p,q}^{\text{intra}} = \sum_{q'=1}^q \hat{\text{ch}}_{\mathcal{A}(p),q'} \quad \text{for all } p = 1 \dots N_{\mathcal{P}}, \quad q = 1 \dots N_{\mathcal{Q}} \quad (5.16)$$

$$\text{gen}_{pq} \in [0, \text{cap}^g] \quad \text{for all } p = 1 \dots N_{\mathcal{P}}, \quad q = 1 \dots N_{\mathcal{Q}} \quad (5.17)$$

$$\text{sto}_p^{\text{inter}} + \text{sto}_{p,q}^{\text{intra}} \in [0, \text{cap}^s] \quad \text{for all } p = 1 \dots N_{\mathcal{P}}, \quad q = 1 \dots N_{\mathcal{Q}}. \quad (5.18)$$

These specify the same power generation balance (5.13), empty initial storage (5.14), storage level changes within and between (*linking*) periods (5.15)-(5.16), and the generation and storage bounds (5.17)-(5.18).

The new problem is symmetric (in the constraints and objective function) in any two generation levels mapped to the same representative period — i.e. any $\text{gen}_{p_1,q}$ and $\text{gen}_{p_2,q}$ for which $\mathcal{A}(p_1) = \mathcal{A}(p_2)$. The same applies to intra-period storage levels $\text{sto}_{p,q}^{\text{intra}}$. Their values are hence equal and we may (without any loss of generality) use the mappings

$$\text{gen}_{p,q} \mapsto \hat{\text{gen}}_{\mathcal{A}(p),q} \quad \text{for all } p = 1 \dots N_{\mathcal{P}}, \quad q = 1 \dots N_{\mathcal{Q}} \quad (5.19)$$

$$\text{sto}_{p,q}^{\text{intra}} \mapsto \hat{\text{sto}}_{\mathcal{A}(p),q}^{\text{intra}} \quad \text{for all } p = 1 \dots N_{\mathcal{P}}, \quad q = 1 \dots N_{\mathcal{Q}}. \quad (5.20)$$

These reduce the number of decision variables from $N_{\mathcal{P}}N_{\mathcal{Q}}$ to $N_{\mathcal{P}_r}N_{\mathcal{Q}}$ each, since each has one unique equation per *representative* period instead of per period. They also reduce the number of constraints for (5.13), (5.16) and (5.17) to $N_{\mathcal{P}_r}N_{\mathcal{Q}}$ each.

Unlike the case without storage (Section 4.1.2.2), we aggregate not only input time series, but also storage (dis)charge decisions. Aggregating only the former fails to make the problem symmetrical in the $\text{gen}_{p,q}$ since the $\text{ch}_{p,q}$ may differ due to (5.16) and (5.18). The extra aggregation comes at a cost; it reduces the problem’s operational flexibility, since storage (dis)charge decisions must be equal across replications of the same representative period. Furthermore, inter-period storage levels $\text{sto}_p^{\text{inter}}$ and storage constraints (5.15) and (5.18) do not reduce; they still appear for each period in the full time series.

In summary, to reduce computational cost in models with storage, we aggregate both time series inputs and storage (dis)charge decisions. Decision variables and constraints then lie on two temporal scales:

- Generation, transmission (in models with multiple regions), storage charging and intra-period storage levels appear once per *representative* period.
- Inter-period storage level constraints appear once per period in the *original* time series.

5.1.3 Literature review

In this section, we expand the literature review in Section 4.1.3 to planning problems with inter-period *linking* constraints such as storage. For additional detail, see [Gonzato et al., 2021].

One approach is to aggregate periods only when they are adjacent chronologically. [Pineda and Morales, 2018] do this to create representative periods with differing lengths. [Tso et al., 2020] merge days, while [De Guibert et al., 2020] increase temporal resolution at times of high net demand. These approaches preserve chronology and may require checking storage levels at fewer points (e.g. the edges of periods), but have reduced representative power since similar events can be aggregated only if one occurs immediately after the other.

Another method is to create representative periods for different regions of the time series. [Welsch et al., 2012] create eight days (one week- and weekend-day for each of the four seasons) and repeat these to recreate a full year. [Samsatli and Samsatli, 2015] do the same, but allow variable time step lengths inside the day. [Timmerman et al., 2017] extend this approach to storage technologies that lose charge over time.

Methods without restrictions on which periods can be aggregated have high representative power but require care to maintain chronology. This is usually done

by *linking* storage levels between representative periods in the order in which they appear in the full time series [Gabrielli et al., 2018]. Storage (dis)charge decisions are equal in each replication of the same representative period, but their absolute storage levels may differ across the full time series. [Tejada-Arango et al., 2018] introduce methods to link both representative periods and individual time steps (referred to as *system states*) using a transition matrix. [Kotzur et al., 2018b] decompose storage levels into the sum of *inter*- and *intra*- period states and consider them on different time scales (discussed in more detail in Section 5.1.4). [van der Heijde et al., 2019] determine the sequence of ordered representative days by minimising a pointwise error metric.

5.1.4 Conclusions and implications

Time series aggregation with inter-period storage involves two sets of assumptions, as depicted in Figures 5.1-5.3. The following are equal in each replication of the same representative period:

1. Demand and weather time series (*time series input aggregation*)
2. Storage (dis)charge decisions (*storage decision aggregation*).

Both approximations induce errors. *Time series input aggregation* reduces the number of unique periods (Figure 5.2), reducing variability. The truncation of extremes especially leads to significant errors in design estimates (Section 4.1.4). *Storage decision aggregation* induces errors due to a reduction in operational flexibility; storage decision variables are constrained to be equal in each replication of the same representative period (Figure 5.3).

There is a mismatch between how representative periods are *determined* and *applied* (Figures 5.1-5.3). Representative periods are *determined* only from input (demand and weather) time series, but aggregation is *applied* also to storage (dis)charge decisions. For this reason, periods with very different storage patterns in the ground truth (non-aggregated) case may be constrained to be equal if they have similar demand and weather patterns.

Aggregation generates decision variables on two temporal scales. Generation, transmission and storage *power* (MW) levels appear once per representative period. Conversely, storage *energy* levels (MWh) are checked for the full time series by recreating the sequence of *ordered* representative periods (Figure 5.3). [Kotzur et al., 2018b] exploit these multiple scales by decomposing storage levels into *inter-period* (level at start of period) and *intra-period* (deviation from start of period, equal in

each replication) contributions. For example, a year of hourly time steps aggregated to ten representative days has 365 inter-day and $240=10\times 24$ intra-day storage levels.

5.2 Contribution

In this chapter, we introduce *storage importance subsampling*: a generalisation of *importance subsampling* to models with storage. It allows us to estimate cost-optimal system design across long samples at reduced computational cost and with lower levels of errors than established schemes. This makes the consideration of multi-year demand and weather samples in storage planning problems, essential to e.g. analysis of future climate, possible. A simulation study demonstrates the method’s reduced errors when compared to previous approaches.

5.3 Methods

In this section, we introduce our method, which we call *storage importance subsampling*. We present an overview in Section 5.3.1, the precise algorithm in Section 5.3.2 and additional remarks in Section 5.3.3.

5.3.1 Overview and intuition

Storage importance subsampling is identical to the storage-free *importance subsampling* apart from two modifications. The first is that, in addition to time series inputs, we use storage (dis)charge decisions when aggregating. As discussed in Section 5.1.4, *a priori* storage aggregation can constrain periods with different storage patterns (in the non-aggregated case) to be equal. In our method, we include (dis)charge decisions in the vectors used to determine representative periods; aggregation is then *determined from* and *applied to* the same variables. The second change is our use of *ordered* representative days and the decomposition into *inter-* and *intra-*period storage levels introduced by Kotzur et al. (2018b) to preserve chronology (see Figure 5.3).

With *storage importance subsampling*, we attempt to reduce aggregation errors via three innovations, as summarised in Table 5.1. We reduce *time series input aggregation* errors by (1) using the medoid as the representative period instead of the mean and (2) including relevant extremes, identified using the *importance* function as in the previous chapter. We reduce *storage decision aggregation* errors by

(3) using storage decisions when aggregating. Methods (2) and (3) rely on operational variables — the *importance* function depends on e.g. generation levels, while storage decisions are themselves operational variables. We calculate these using a preliminary planning and operation model run as in the previous chapter.

5.3.2 Algorithm: *storage importance subsampling*

Inputs:

- $(\xi_p)_{p \in \mathcal{P}}$: demand and weather time series, length $n_{\mathcal{P}}$ periods
- $n_{\mathcal{A}}$: number of unique representative periods to aggregate into
- $p_e \in [0, 1]$: proportion of periods in $\mathcal{P}$ considered *extreme*
- IMP: real-valued *importance* function, from operational variables $\mathbf{O}_p$

Steps:

1. Get preliminary optimal design estimate $\hat{\mathbf{D}}_{\mathcal{A}_0}$:
 - (a) Aggregate $(\xi_p)_{p \in \mathcal{P}}$ into $(\hat{\xi}_{\mathcal{A}_0(p)})_{p \in \mathcal{P}}$, with $n_{\mathcal{A}}$ unique representative periods, using time series inputs.

- (b) Solve planning problem:

$$\hat{\mathbf{D}}_{\mathcal{A}_0} = \text{PSM}_{\text{plan}}((\hat{\xi}_{\mathcal{A}_0(p)})_{p \in \mathcal{P}}). \quad (5.21)$$

2. Create *importance subsample*:

- (a) Determine system operation across full time series:

$$(\hat{\mathbf{O}}_p)_{p \in \mathcal{P}} = \text{PSM}_{\text{operate}}((\xi_p)_{p \in \mathcal{P}} \mid \hat{\mathbf{D}}_{\mathcal{A}_0}). \quad (5.22)$$

- (b) Calculate *importance* of each period:

$$\text{imp}_p = \text{IMP}(\hat{\mathbf{O}}_p) \quad \text{for all } p \in \mathcal{P}. \quad (5.23)$$

- (c) Partition time series $\mathcal{T}$ into:

- $\mathcal{P}_e$: $p_e n_{\mathcal{P}}$ *extreme* periods (with highest *importance*)
- $\mathcal{P}_r$: $(1 - p_e) n_{\mathcal{P}}$ *regular* periods (those remaining).

- (d) Create *importance subsample* $(\hat{\xi}_{\mathcal{A}_1(p)})_{p \in \mathcal{P}}$, $n_{\mathcal{A}}$ periods, with:

- $\frac{n_{\mathcal{A}}}{2}$ *extreme* representative periods aggregated from $\mathcal{T}_e$
- $\frac{n_{\mathcal{A}}}{2}$ *regular* representative periods aggregated from $\mathcal{T}_r$.

Aggregate using time series values and intra-day storage levels.

3. Get final optimal design estimate $\hat{\mathbf{D}}_{\mathcal{A}_1}$:

$$\hat{\mathbf{D}}_{\mathcal{A}_1} = \text{PSM}_{\text{plan}}((\hat{\xi}_{\mathcal{A}_1(p)})_{p \in \mathcal{P}}). \quad (5.24)$$

Output: $\hat{\mathbf{D}}_{\mathcal{A}_1}$: optimal design estimate

Source of error	Impact on aggregated data	Approach to mitigate
Time series input aggregation		
(1) Representative period is cluster mean	Lower variability	Use cluster medoid
(2) Extremes truncated	Contains less intense extreme events	Identify and include relevant extremes
Storage decision aggregation		
(3) Aggregation ignores storage patterns	Periods with differing storage patterns constrained to be equal	Use storage patterns in aggregation

Table 5.1: How *storage importance subsampling* reduces aggregation error.

5.3.3 Remarks

In this section, we provide additional information and implementation details for the algorithm. The remarks for the storage-free version (Section 4.3.3) still apply, but we do not repeat them here.

In the planning runs in Steps 1(b) and 3, we use *ordered representative periods* as discussed in Section 5.3.1. Since storage levels have to be checked across the full time series, computational cost increases with the length of the full time series even with a constant number of representative periods. However, we find this effect to be relatively modest: a simulation with 120 representative days from five years takes roughly 10% longer than from three years.

Storage importance subsampling depends on a number of hyperparameters. Like in Chapter 4, the real-valued *importance* function identifies extreme events; we use the generation cost in our case study. The role of p_e is slightly different as seen by comparing Figures 4.3 and 5.4. For the storage-free *importance subsampling*, the aggregated sample includes the $p_e n_A$ periods with the highest *importance* in the full time series. *Storage importance subsampling* instead aggregates the $p_e n_P$ periods with highest importance into $\frac{n_A}{2}$ representative ones. This is for a number of reasons. For models without storage, capturing the instantaneous peak demand (after subtracting renewables) is essential (since this drives capacity), so we select *all* the highest values to avoid peak truncation. In contrast, storage is usually used for peak shaving, meaning capturing exact peaks is less important. In this case, we opt to denote a larger portion of the full time series as *extreme* — $p_e n_P$, a fraction of the *full* time series, is larger than $p_e n_A$, a fraction of the *aggregated* time series — and represent it at higher resolution. For simplicity, we use equal numbers

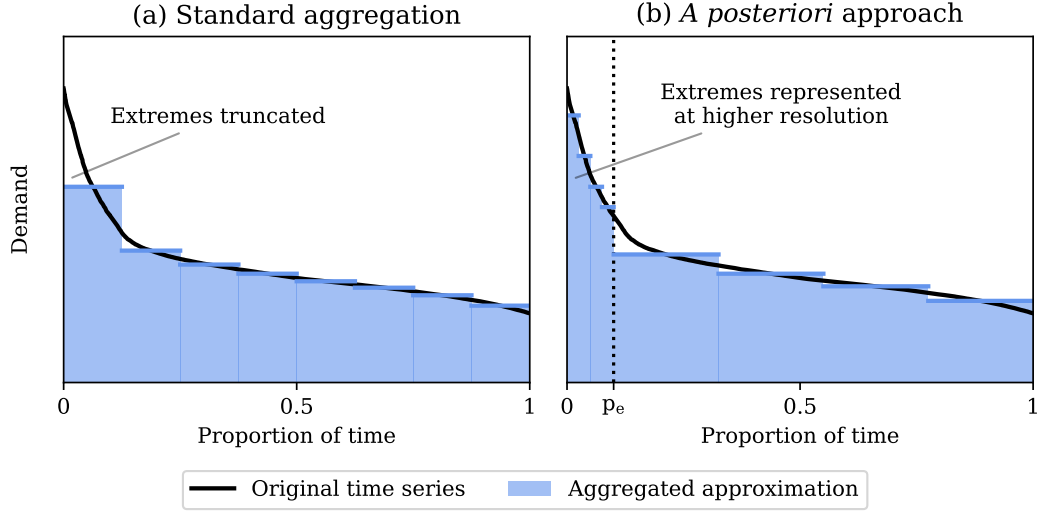


Figure 5.4: Impacts of time series aggregation. Black line: *duration curve* of the *importance* (e.g. generation cost or electricity price), which identifies extreme events. (a) Standard aggregation truncates extremes with a high *importance*. (b) *Storage importance subsampling* represents extremes at higher resolution.

of representative periods $\frac{n_A}{2}$ for both the *extreme* and *regular* regions instead of introducing a new hyperparameter for this split.

5.4 Simulation studies

5.4.1 Overview

In this section we evaluate *storage importance subsampling*'s performance on the *Continuous Linear Program with Storage (CLPwS)* model introduced in Chapter 3. As in the previous chapter, we compare optimal system design estimates with the actual optima across the full time series. We also calculate unmet demand levels across the full time series. Unlike the previous chapter, we do not consider *peak capacity shortage* (maximum demand minus available generation capacity) since this is not well-defined; storage can displace unmet demand in time, and we can have high unmet demand for a short time or lower shortfalls over a longer period.

We repeat each experiment 40 times with different base time series $(\xi_p)_{p \in \mathcal{P}}$ created by resampling years with replacement; for example, a three-year base time series may be [2011][1992][1992]. We consider the distribution of each technology's normalised capacity error, defined as $\frac{\text{cap}^A - \text{cap}^T}{\text{cap}^T}$ where cap^A and cap^T are capacities across aggregated and non-aggregated data respectively. Like Chapter 4.1, we examine only systemwide totals, since many capacities at regional level are zero,

making capacity errors undefined.

We investigate each of *storage importance subsampling*'s three innovations sequentially. We run four aggregation schemes:

- A. cluster on time series inputs, use mean as representative (*a priori*)
- B. cluster on time series inputs, use medoid as representative (*a priori*)
- C. identify and include extremes using *importance* function, cluster on time series inputs (*a posteriori*)
- D. *storage importance subsampling*: identify and include extremes using *importance* function, cluster on time series inputs and storage decisions (*a posteriori*).

5.4.2 Setup

Simulation details are provided in Table 5.2. We run two sets of simulations, in which we aggregate three years into 30 and 120 representative days respectively. We use daily periods with 24 hourly time steps each. Whenever we cluster (in methods A and B, as well as in Steps 1(a) and 2(d) for methods C and D), we normalise each time series to mean zero and variance one using a z -transform, reshape them to daily vectors and cluster these using Ward's hierarchical clustering. We also ran simulations in which we normalise time series to the range $[0, 1]$ or use iterative (e.g. k -means) clustering, with either worse or similar performance.

Under aggregation, the number of decision variables changes as follows. All problems have 19 design decision variables: baseload, peaking, wind and storage energy capacities in three regions each, and seven transmission capacities. A non-aggregated problem has 22 operational decision variables per time step: 19 (one for each design variable) generation, transmission and storage (dis)charging levels, and three storage energy levels¹. This makes $N_{\mathcal{P}} \times 24 \times 22$ across $N_{\mathcal{P}}$ days with hourly time steps. The same problem aggregated into $N_{\mathcal{P}_r}$ representative days has $N_{\mathcal{P}_r} \times 24 \times 22$ operational variables (the same 22 as before, one set per *representative* period; storage energy levels are replaced by their *intra-period* counterparts as in Section 5.1.2) as well as $N_{\mathcal{P}}$ inter-period storage energy levels (one per period in the full time series). For example, a non-aggregated three year problem has 19 design and $1095 \times 24 \times 22 = 578160$ operational decision variables. The same problem

¹These are not actually independent (they are rolling sums of (dis)charging levels) but we separate them for clarity.

Aggregation	Run	Number of (repr.) days	Solution time [min]	
			Mean	95% of runs
No aggregation (target)				
	None	Total	1095*	644 (313 - 1568)
30 representative days				
A: <i>A priori</i> , mean representative	Total	30	2	(1 - 4)
B: <i>A priori</i> , medoid representative	Total	30	3	(1 - 5)
C: <i>A posteriori</i> , with extremes, aggregate on time series inputs	Total	—	94	(14 - 202)
	P1	30	3	(1 - 5)
	O	1095*	89	(10 - 197)
	P2	30	3	(2 - 6)
D: <i>A posteriori</i> , storage importance subsampling	Total	—	94	(14 - 204)
	P1	30	3	(1 - 5)
	O	1095*	89	(10 - 197)
	P2	30	3	(2 - 5)
120 representative days				
A: <i>A priori</i> , mean representative	Total	120	302	(84 - 808)
B: <i>A priori</i> , medoid representative	Total	30	250	(49 - 524)
C: <i>A posteriori</i> , with extremes, aggregate on time series inputs	Total	—	562	(204 - 868)
	P1	120	250	(49 - 524)
	O	1095*	218	(21 - 465)
	P2	120	206	(63 - 708)
D: <i>A posteriori</i> , storage importance subsampling	Total	—	573	(273 - 996)
	P1	120	250	(49 - 524)
	O	1095*	218	(21 - 465)
	P2	120	215	(66 - 643)

* 1095-1098 days depending on leap days

Table 5.2: Simulations in Section 5.4. For *a posteriori* methods, we disaggregate into first planning (P1), operational (O) and second planning (P2) run, corresponding to Steps 1(b), 2(a) and 3 respectively in Section 5.3.2.

aggregated into 30 days has 19 design variables and $30 \times 24 \times 22 + 1095 = 16935$ operational decision variables.

For method A, we cluster the daily vectors of the six input time series (three demand and three wind, giving each vector $6 \times 24 = 144$ components) and use the cluster mean as the representative. Method B is identical except with the *medoid* (closest daily vector to cluster mean in normalised space) as the representative.

Method C is *storage importance subsampling*, except that in Step 2(d) we aggregated using only time series values and not intra-day storage levels. Method D *does* use this information, and hence aggregates vectors with $9 \times 24 = 216$ components each, which contain $6 \times 24 = 144$ components for the time series inputs and $3 \times 24 = 72$ components for inter-day storage levels.

For methods C and D, we use $p_e = 0.05$; 5% of the original time series is considered *extreme* and represented at higher resolution. For the *importance* function, we use daily total generation cost as in Chapter 4. In steps 1(a) and 2(d), we use the medoid as the representative day.

5.4.3 Results

Figure 5.5 shows the distribution of capacity errors and unmet demand across the test simulations. Reading plots from left to right (A-D) presents the impact of each of innovation (using the medoid as representative, identifying and including extremes, and using storage patterns in aggregation).

Method A, with *a priori* aggregation and mean representative days, underestimates baseload and transmission while overestimating peaking and wind capacity. Furthermore, we find a drastic underestimation of optimal storage energy capacity with 30 representative days: a median error of 60%, with over 90% in some simulations. These errors lead to large levels of unmet demand: a median of 0.5% and up to almost 1.5%. In financial terms, a value of lost load of £6,000/MWh, as in Section 3.5.1.1, implies an increase of roughly 1% in system cost for every 0.01% of unmet demand.

Method B, which uses the medoid as the representative, shows considerably better performance. For baseload, peaking, wind and transmission capacities, the biases are small — typically less than 5% with 30 representative days. However, storage is still underestimated significantly (60% and 20% for 30 and 120 representative days respectively). Unmet demand is less than half of method A.

Identifying and including relevant extreme events in method C increases optimal

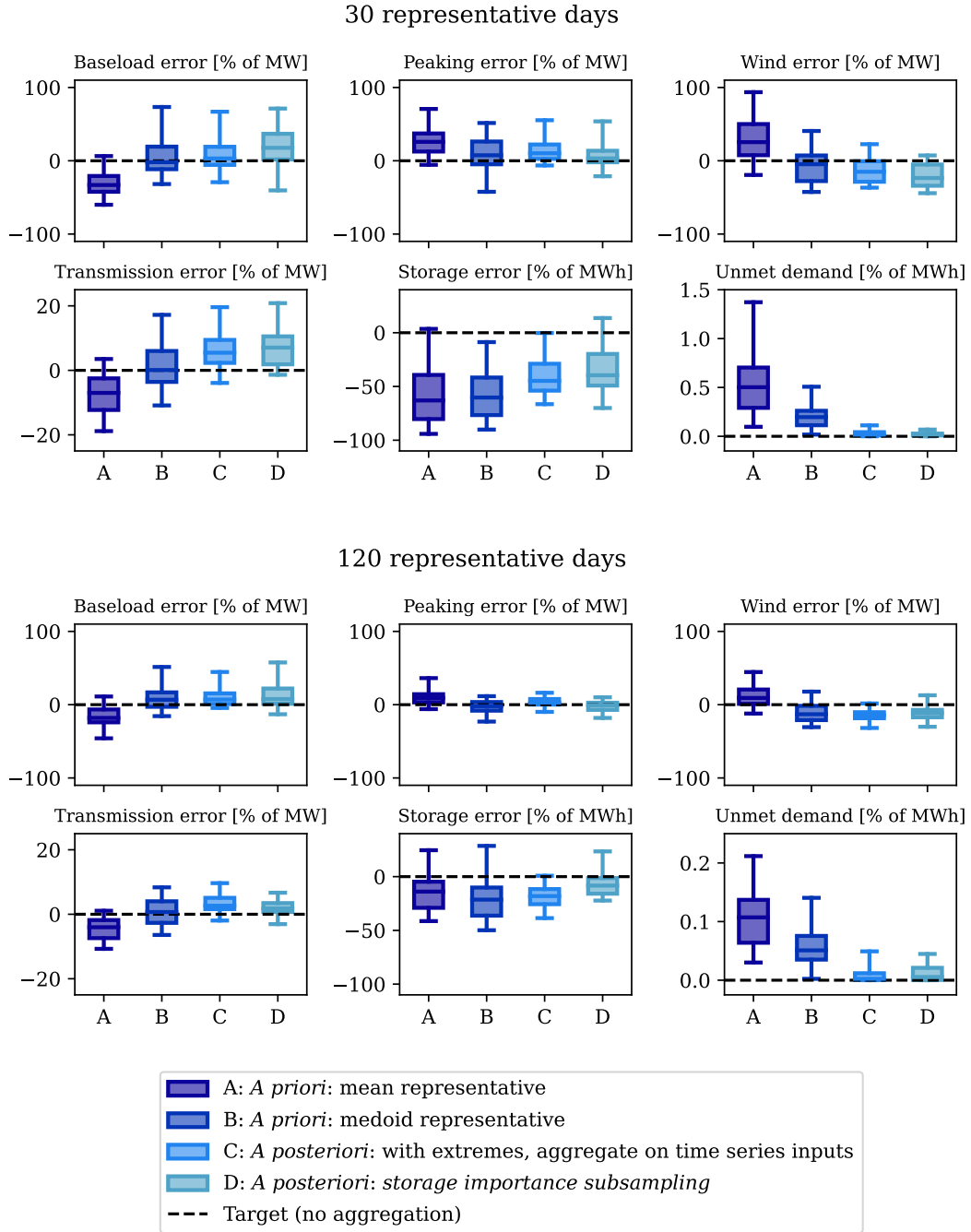


Figure 5.5: Distribution of capacity errors and unmet demand across 40 aggregated samples from methods A-D from Section 5.4.1. Box-and-whiskers: 2.5%, 25%, median, 75% and 97.5% percentiles. Dashed black line: optimum across full time series (target). Unmet demand is expressed as percentage of total demand.

capacity error slightly (with wind slightly underestimated and transmission slightly overestimated), but unmet demand levels are reduced to almost zero. This is to be expected, since the extreme events force the planning model to meet demand in such scenarios.

Storage importance subsampling, which uses storage information in aggregation, reduces the underestimation of storage capacity: 40% and 5% for 30 and 120 representative days respectively. Impacts on other capacity errors are minimal. Levels of unmet demand are slightly higher than method C but remain small.

For all methods, increasing the number of representative days from 30 to 120 decreases error metrics. For example, *a priori* aggregation with method B has lower capacity error and unmet levels with 120 representative days than *storage importance subsampling* has with 30.

5.4.4 Discussion

The performance of methods A-D illustrate the impacts of *storage importance subsampling*'s innovations. Using the medoid instead of the mean as representative reduces most error metrics by at least 50%. This is in agreement with previous studies (Section 4.1.4), which also recommend the use of closest-day based aggregation. This is an *a priori* approach which does not necessitate any operational variables or preliminary model runs. Hence, we recommend always using medoids as representative periods.

Improvements from the *a posteriori* modifications differ. Like previous studies (Section 4.1.4), identifying and including extreme events (method C) leads to design estimates that are more robust to such occurrences. The performance gain from using storage information (method D, the full *storage importance subsampling*) is concentrated primarily in more accurate estimates of optimal storage capacity. In fact, this approach leads to slightly higher levels of unmet demand (although still very low). Hence, using storage information in aggregation (at least in the way we do, by concatenating intra-day storage levels to the cluster vectors) may not be required in all modelling contexts.

In general, estimating optimal storage energy capacity appears to be difficult, with models underestimating it by large margins. This may be because storage energy capacities are driven by rolling sums of demand shortfalls instead of instantaneous values; a small error repeated ten times becomes ten times as large.

Similarly to Chapter 4, solution times are distorted by the requirement for a

benchmark run with no aggregation. We could, for example, estimate a ten-year optimal design using 120 representative days using *storage importance subsampling*; we do not do this because we are unable to verify this against the ground truth. With longer simulations, the two planning runs would dominate solution times over the operational run. Hence, as in the previous chapter, *storage importance subsampling* requires roughly twice the computational expense as an *a priori* scheme.

5.5 Conclusions

This chapter’s contribution is *storage importance subsampling*, a time series aggregation scheme for power system planning models with storage. It combines two previous classes of aggregation methods: those that preserve chronology and hence work with storage technologies, and *a posteriori* methods that adapt the aggregation to the underlying planning model.

The implications are similar as Chapter 4: our method allows the consideration of long samples in power system planning models with lower levels of error than established approaches. This opens the door to the use of climate model data in such studies.

We see a number of natural extensions. As in the previous chapter, we can investigate different *importance* functions than the generation cost. We may also use different operational variables when aggregating in Step 2(d) (Section 5.3.2). In our implementation, we concatenate the vector of *intra-day* storage levels to the demand and weather data. However, we could also use e.g. the storage (dis)charge decisions, or other operational variables (such as using the generation cost, currently used only used to stratify). As discussed in Section 4.5, we can also iterate the process multiple times, alternating between planning and operational runs. We did this in our case study but did not find any additional performance gains. Other possible changes include a different approach to preserve chronology than the linking of representative periods as we do currently, such as the *system states* or *chronological time period clustering* methods discussed in Section 5.1.3. Finally, dimensionality reduction may be required for models with more time series. In our case study, we have demand, wind and storage at three locations each. This generates daily vectors with $3 \times 3 \times 24 = 216$ components. With more regions and time series, these vectors may become impractically long.

Our use of planning models with perfect operational foresight — we optimise operation with knowledge of the demand and weather data in each time step — is

not realistic. For models without storage, operation is essentially independent across periods; knowledge of the future does not affect operational decisions. For storage, however, this is not the case, and allowing storage optimisation with knowledge of the long-term future is unrealistic. An interesting generalisation is to implement operational policies without perfect foresight.

6. Efficient uncertainty quantification

In this section, we introduce a subsampling algorithm that efficiently quantifies demand and weather uncertainty on power system planning model outputs. The method is an m -out-of- n bootstrap modified for time series data. The ideas in this chapter have been published in the following paper:

- [\[Hilbers et al., 2021\]](#). AP Hilbers, DJ Brayshaw, A Gandy. Efficient quantification of the impact of demand and weather uncertainty in power system models. *IEEE Transactions on Power Systems*.

Sample code is available at doi:[10.5281/zenodo.3840911](https://doi.org/10.5281/zenodo.3840911).

This chapter is structured as follows. Section [6.1](#) introduces the problem and established approaches. Section [6.2](#) describes our contribution. Section [6.3](#) introduces the method and Section [6.4](#) provides a case study of its application on the planning models of Chapter [3](#). Section [6.5](#) makes concluding remarks.

6.1 Background

6.1.1 Context: uncertainty quantification

Recall from Section [2.3.3](#) that planning models may be viewed as a mapping from demand and weather data $(\xi_t)_{t \in \mathcal{T}}$ to the associated optimal design $\hat{\mathbf{D}}$:

$$\hat{\mathbf{D}} = \text{PSM}_{\text{plan}}((\xi_t)_{t \in \mathcal{T}}). \quad (6.1)$$

Sampling uncertainty on $\hat{\mathbf{D}}$ is induced from the demand and weather time series $(\xi_t)_{t \in \mathcal{T}}$ used to calculate it; a different sample, drawn from the same underlying distribution, leads to a different value. Section [2.5](#) indicates that this *demand and weather uncertainty* can be large. In this chapter, we quantify it efficiently in both data and computation. We consider only this uncertainty and not those in e.g. costs, technology availabilities and policy.

Users might not actually seek $\hat{\mathbf{D}}$, but $\mathbf{D}$, the solution to the stochastic programming problem (Section [2.7](#)) that minimises *expected* costs under demand and weather sampling uncertainty. However, this is usually unavailable, as it requires knowledge

of the underlying distribution. Users may estimate it via $\hat{\mathbf{D}}$, from a planning model (the *sample average approximation*, see Section 2.7). Quantifying uncertainty on $\hat{\mathbf{D}}$ — which comes from a deterministic problem given uncertain data — is hence important, since it gives an indication whether design estimates are statistically robust or a product of the particular demand and weather sample. Furthermore, it may provide information on $\mathbf{D}$ via e.g. confidence intervals, which depend on the sampling distribution of $\hat{\mathbf{D}}$.

6.1.2 Literature review

The power system modelling community employs a number of established techniques to quantify output uncertainty, as summarised by Yue et al. [2018]. We describe a number of approaches and comment on their relevance below.

6.1.2.1 Scenarios

Scenario analysis involves running a power system model on different sets of assumptions picked by users to represent possible circumstances. For example, the UK’s Committee on Climate Change [2020] models the UK power system under a number of possible future electricity demand levels, driven by different electrification costs and adoption rates. Similarly, Jenkins et al. [2021] determine US energy system design in cases without restrictions on technological buildout versus ones where nuclear power is prohibited or renewable capacity install rates are constrained.

Scenario analysis is often required for factors that cannot reliably be described by probability distributions, such as future policy, technological and cost developments, or human behaviour. One set of model outputs is generated for each scenario, without probabilities. This restricts uncertainty quantification; statements like “in this scenario, 100GW of wind capacity is cost-optimal” are obtainable, whereas statements like “there is a 70% chance the cost-optimal level of wind capacity is at least 80GW” are usually not.

6.1.2.2 Sampling-based methods

Quantitative information on input values allows model output variability to be quantified. For example, *interval analysis* runs a model with the minimum and maximum value of each input parameter [Soroudi and Amraee, 2013] to obtain bounds on outputs. *Monte Carlo* methods sample each input parameter according to its probability distribution; the empirical distribution of model outputs approximates

the output’s probability distribution as described in Appendix [A.1.2](#). Monte Carlo methods have been applied in many power system applications; we discuss a selection here. [Alzbutas and Norvaisa \[2012\]](#) use uncertainties in install and generation costs to estimate the probability distribution of a future Lithuanian power system’s costs. [Pye et al. \[2015\]](#) conduct a similar analysis with a larger set of uncertain parameters for a UK system. [Bosetti et al. \[2015\]](#) examine the sensitivity of e.g. proportions of renewable-generated electricity to various input costs.

The applicability of interval and Monte Carlo analysis differs per setting. Interval analysis requires a minimum and maximum value of each input parameter to be known or assumed, and the extrema of these inputs only translate to extrema of model outputs if each input-output relationship is monotonic. It also fails to provide probabilities and suffers from a *curse of dimensionality*; the number of model runs is $2^{\text{[number of inputs]}}$. In contrast, Monte Carlo methods can be used to estimate probability distributions of model outputs, and can do so efficiently, since many uncertain inputs can be sampled at once to generate e.g. 1000 samples (see Appendix [A.1](#)). However, they require each inputs’ probability distribution to be known.

6.1.2.3 Emulators

Another approach is to create an *emulator* — a statistical approximation (usually much faster to evaluate) of the model’s input-output relationship. This is *trained* on a number of known input-output pairs, often generated by running the original model. For example, [Lawson et al. \[2016\]](#) use a Gaussian process trained on a number of judiciously chosen inputs, with associated outputs coming from a UK transmission expansion problem. The emulator then efficiently estimates model outputs, as well as uncertainty bounds (which a Gaussian Process provides), for any inputs.

Since emulators are usually faster to evaluate than the functions they approximate, they allow uncertainty analysis on a larger set of inputs than the original model. However, the emulator must be trained, requiring additional work, both computationally and for the user. Furthermore, outputs from the emulator away from the training data may be inaccurate, as they rely on interpolating or extrapolating to new values.

6.1.2.4 Bootstrap methods

Bootstrap methods, as introduced by [Efron \[1979\]](#), can quantify sampling uncertainty. In their most basic form, they approximate an estimator’s sampling distri-

bution by its distribution under resampling with replacement from available data. They provide consistent uncertainty estimates for a wide variety of estimators — such as sample means, quantiles, and maximum likelihood estimators — with minimal modifications and without individual limit laws [Bickel and Freedman, 1981]. They do not improve point estimates, but can indicate an estimator’s variability, estimating variance, probabilities or confidence/prediction intervals. For the machinery, see Appendix A.3.1.

Bootstrap methods can be modified for special settings. For example, the *m*-out-of-*n* bootstrap, which calculates an estimator’s sampling distribution across resamples of a different length (usually shorter) than the original dataset, may be used to reduce computational cost when calculating the estimator’s value is computationally expensive. See Appendix A.3.2 for theory, including consistency and sample length rescaling requirements. Another generalisation is for time series where values are not independent and identically distributed (IID); two common modifications are the sampling of blocks of time steps to preserve short-term dependence structures (the *block bootstrap*) and detrending of long-term dependencies and subsequent resampling of residuals (the *model-based bootstrap*) [Davison and Hinkley, 1997, Chapter 8].

Bootstrap methods have been applied in a number of power system applications. [Fan and Hyndman, 2012] use a model-based block bootstrap (resampling model residuals across blocks of time steps) to create probabilistic forecasts of electricity demand, training a new model for each sample. [Chen et al., 2012] and [Wan et al., 2014] predict electricity price and wind generation respectively, allowing the forecast’s variance to depend on the dependent variables. [Khosravi et al., 2013] resample a sliding window of observations, a procedure called the *moving block bootstrap*, for probabilistic forecasts of wind generation. [bin Othman et al., 2008] use a model-based bootstrap for a probabilistic assessment of system adequacy margins.

6.1.3 Conclusions and implications

Not all uncertainty analysis methods introduced above can quantify the demand and weather uncertainty on $\hat{\mathbf{D}}$ in (6.1). Both scenario and interval analysis are impractical due to the high dimensionality of the input time series (multiple time series and many time steps) and the fact that such approaches provide only a range of possible outcomes, but no probabilities. Standard Monte Carlo methods require many independent samples, while we have only a single time series of data. Bootstrap

methods *are* possible, since we can run models across artificial samples created from resampling. This must be done carefully to preserve temporal and spatial dependence structures. We introduce such a scheme in this chapter.

6.2 Contribution

In this section, we provide two contributions. The first is an uncertainty quantification scheme for design estimates from power system planning models under demand and weather uncertainty. It is an m -out-of- n bootstrap using *shorter* re-samples of the available data. The second is a method to determine the required simulation length for desired uncertainty bounds. We validate our methods experimentally and introduce a diagnostic for the consistency of uncertainty estimates that is found to work well.

We illustrate improvements on currently employed approaches (Section 6.1.2.2) by an example. Consider uncertainty quantification for a design estimate obtained using one year of demand and weather data. The standard Monte Carlo approach is to run the model multiple times with different (disjoint) one-year samples. This is inefficient, both in data (many year-long samples) and computation (many one-year PSM runs). Our method conducts simulations across *shorter* datasets (with reduced computational cost) generated from what is already available (hence requiring no additional data).

6.3 Methods

6.3.1 Overview

We consider each scalar component $\hat{D}$ of $\hat{\mathbf{D}}$ separately. One quantification of the demand and weather uncertainty on $\hat{D}$ is provided by its sampling standard deviation $\sigma_{\hat{D}}$ across samples $(\xi_t)_{t \in \mathcal{T}}$ in (6.1). $\sigma_{\hat{D}}$ implies confidence or prediction intervals; for example, $[\hat{D} - 2\sigma_{\hat{D}}, \hat{D} + 2\sigma_{\hat{D}}]$ has a coverage probability of at least 75% (and usually higher, e.g. 95% for a normal distribution).

As discussed in Section 6.2, we often cannot estimate $\sigma_{\hat{D}}$ with disjoint samples due to limitations in data and computation. We estimate it using the m -out-of- n bootstrap below.

6.3.2 Algorithm: m -out-of- n bootstrap for $\sigma_{\hat{D}}$

Inputs:

- $(\xi_t)_{t \in \mathcal{T}}$: demand and weather data used for $\hat{D}$, $n = |\mathcal{T}|$ periods
- m : number of periods in subsample
- K : number of subsamples

Steps:

1. Construct K subsamples $\mathcal{S}_1, \dots, \mathcal{S}_K \subset \mathcal{T}$, of length $m \leq n$, using a suitable bootstrapping procedure (see Section 6.3.3).
2. Calculate $\hat{D}_k = \text{PSM}_{\text{plan}}((\xi_t)_{t \in \mathcal{S}_k})$ for each sample $k = 1, \dots, K$.
3. Estimate variance $\widehat{\sigma_S}^2$ across subsamples:

$$\widehat{\sigma_S}^2 = \frac{1}{K-1} \sum_{k=1}^K (\hat{D}_k - \bar{z_S})^2 \quad (6.2)$$

where $\bar{z_S} = \frac{1}{K} \sum_{k=1}^K \hat{D}_k$ is the sample mean.

4. Estimate $\sigma_{\hat{D}}$ by $\widehat{\sigma_{\hat{D}}}$, defined by

$$\widehat{\sigma_{\hat{D}}}^2 = \frac{m}{n} \widehat{\sigma_S}^2. \quad (6.3)$$

Output: $\widehat{\sigma_{\hat{D}}}$: estimate of $\sigma_{\hat{D}}$

6.3.3 Remarks

The procedure in Step 1 works by subsampling, with replacement, blocks of time steps correctly distributed throughout the year. An example is the sampling of weeks from each season. Precise details (e.g. the block length) may be tailored to the specific application as in Section 6.4.1.

The relationship between sample length and standard deviation allows the estimation of the required point estimate sample length n to attain a desired precision. Given a required standard deviation $\sigma_{\hat{D}}$, rearrange (6.3) to give

$$n = m \frac{\widehat{\sigma_S}^2}{\widehat{\sigma_{\hat{D}}}^2}. \quad (6.4)$$

This motivates the following approach: (1) use Steps 1-3 to obtain $\widehat{\sigma_S}^2$. (2) estimate n , the minimum sample length required to ensure model output $\hat{D}$ has standard deviation at most $\widehat{\sigma_{\hat{D}}}$, using equation (6.4).

6.3.4 Justification and a test for validity

We calculate the standard deviation estimate $\widehat{\sigma}_{\hat{D}}$ using an m -out-of- n time series bootstrap. From bootstrap theory (Appendix A.3), $\widehat{\sigma}_{\hat{D}}$ is consistent for $\sigma_{\hat{D}}$ in many settings if sampled blocks are long enough to capture autocorrelation (so that datasets created by resampling are similar to those obtained from the underlying distribution). As demonstrated in Appendix A.3.2, the $\frac{m}{n}$ scaling appears because the variance of many estimators (including arbitrary functions of the sample mean and median under weak conditions) are inversely proportional to sample size [Bickel et al., 1997, Van der Vaart, 2014].

There are two cases in which the method is *not* expected to provide consistent estimates. As discussed in Appendix A.3, bootstrap consistency relies on two distribution functions (one using the original distribution and one using bootstrap resampling) having the same limit. This is not the case in two settings. The first is when $\hat{D}$ depends on a sample minimum or maximum, e.g. generation capacities if 100% of demand must be met. Sample maxima are often used as canonical examples of bootstrap failure [see e.g. Bickel and Freedman, 1981, counterexample 2], which occurs because, while the distribution of the sample maximum under the underlying distribution is continuous, its distribution under bootstrap resampling has a point mass at the largest value in the sample that is resampled from. The second is when $\hat{D}$ is a discrete variable; in this case distribution functions are discontinuous, invalidating many bootstrap consistency arguments. There are cases where the bootstrap can still provide approximate uncertainty estimates (e.g. when the discrete jumps are small enough to be approximated well by a continuous variable), but the asymptotic results we consider do not apply directly and such behaviour cannot be guaranteed.

A diagnostic gives an approximate indication whether the method provides consistent estimates. It works as follows: check whether the estimate $\widehat{\sigma}_{\hat{D}}$ is (roughly) constant under subsample length m . The diagnostic is necessary but not sufficient; standard deviation estimates are inconsistent if they clearly differ across sample lengths, but constant values do not prove the consistency of the bootstrap. In our case studies, the diagnostic is found to work well, successfully identifying the settings where standard deviation estimates are (in)consistent.

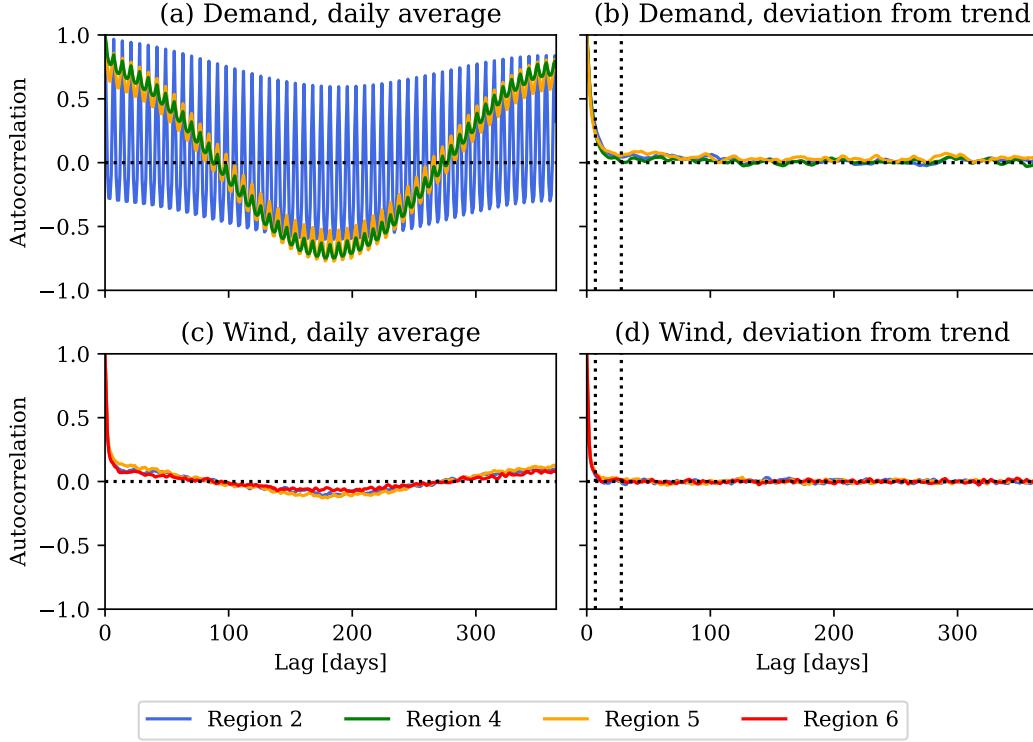


Figure 6.1: Autocorrelation function (ACF) of demand and wind time series. (a) and (c) are ACFs of daily averages while (b) and (d) are ACFs of detrended time series. The dashed vertical lines indicate 7 and 28 days respectively.

6.4 Simulation studies

In this section we apply our method (Section 6.3.2) to the models introduced in Chapter 3. Section 6.4.1 discusses the subsampling scheme used in Step 1. We quantify output uncertainty in Section 6.4.2 and determine the required sample length for desired confidence levels in Section 6.4.3. Section 6.4.4 examines the method's accuracy, while Section 6.4.5 discusses results and their implications.

6.4.1 Setup

We apply the method to the *CLP* and *MILP* models as described in Chapter 3. Generation technologies are modified slightly, with *nuclear*, *combined cycle gas turbine* (CCGT) and *open cycle gas turbine* (OCGT) the dispatchable technologies and wind the only renewable technology. The only difference between the *CLP* and *MILP* models is, in the latter, the ramping and integer constraint (capacity in units of 3GW) for nuclear (see (3.16)-(3.17)). We allow unmet demand at high cost. See

[Hilbers et al., 2021] for details¹.

To avoid distorting power system operation, the K bootstrap samples (Step 1, Section 6.3.1) should preserve temporal dependence structures with respect to the original time series. For demand, we expect a daily, weekly and annual cycle, as well as a short-term autocorrelation; a higher-than-average demand hour is more likely to be followed by another. For wind, we expect an annual and possibly a daily cycle, as well as a short-term autocorrelation.

To understand temporal dependence structures, we plot the time series' autocorrelation functions in Figure 6.1. Figure 6.1(a) and (c) examine the daily mean (hence with daily cycle already removed) demand and wind time series respectively. The weekly and annual cycles are oscillations with periods of seven and 365 days respectively. Their relative strengths vary; in Region 2 (data from Germany), weekly cycles dominate the annual cycles, while Regions 4 and 5 (data from France and the United Kingdom) have stronger seasonal cycles. For wind (Figure 6.1(c)) we find a seasonal cycle, with higher wind speeds in winter.

We expose short-term autocorrelation by removing the weekly and seasonal cycles (by subtracting, from each day, the deviation of that week- and calendar- day from the long-run average) in Figure 6.1(b) and (d). Most short-term dependence is contained within seven days, and almost all within 28 days.

For this reason, we propose two stratified block subsampling schemes:

- *Months*: sample individual months, correctly distributed throughout the year.

For example, a two-year sample is:

$$[\text{Jan}][\text{Feb}] \cdots [\text{Nov}][\text{Dec}][\text{Jan}][\text{Feb}] \cdots [\text{Nov}][\text{Dec}] \quad (6.5)$$

where e.g. $[\text{Jan}]$ is a January block from the original time series $\mathcal{T}$.

- *Weeks*: sample individual weeks, correctly distributed throughout the year.

For example, a 56-day sample is:

$$[7d]_{\text{DJF}}[7d]_{\text{MAM}}[7d]_{\text{JJA}}[7d]_{\text{SON}}[7d]_{\text{DJF}}[7d]_{\text{MAM}}[7d]_{\text{JJA}}[7d]_{\text{SON}} \quad (6.6)$$

where e.g. $[7d]_{\text{DJF}}$ is a week block from the months Dec-Jan-Feb in $\mathcal{T}$.

Under this setup, we preserve daily, weekly and seasonal cycles by design. We also preserve any short-term autocorrelation within each block (one month and seven

¹These modifications were made in response to peer review feedback.

Sampling method	Samples for standard deviation estimate		Mean computing	
	Subsampling scheme	Length m [days]	Time [min]	Memory [MB]
(a) <i>CLP</i> model, point estimate calculated across 2008-2017				
Disjoint Monte Carlo	-	3,650	262	13,813
m -out-of- n bootstrap	<i>months</i>	365	6	1633
(b) <i>MILP</i> model, point estimate calculated across 2017				
Disjoint Monte Carlo	-	365	765	1,713
m -out-of- n bootstrap	<i>weeks</i>	84	14	457

Table 6.1: Simulations in Section 6.4.2. m is the length of samples across which the standard deviation is estimated. *Disjoint Monte Carlo* uses disjoint samples of the same length as the point estimate.

days respectively).

6.4.2 Quantifying output uncertainty

In this section, we quantify the demand and weather uncertainty on design estimate $\hat{D}$, determined using a single long simulation. We seek $\hat{D}$'s sampling standard deviation $\widehat{\sigma}_{\hat{D}}$. We estimate it using the m -out-of- n bootstrap (Section 6.3.1) with $K=1000$ bootstrap samples, drawn from the period 2008-2017. We compare this to standard deviation estimates using disjoint samples of the same length as the point estimate, which we refer to as *disjoint Monte Carlo*. Details of each of these simulations are provided in Table 6.1.

In test case (a), we consider the *CLP* model, with a point estimate calculated across 2008-2017 and the standard deviation estimates using one-year bootstrap resamples, generated by the *months* scheme. In test case (b), the *MILP* model is run across 2017 for the point estimate and standard deviation estimates use subsample length $m=12$ weeks, generated by resampling weeks from seasons as per the *weeks* scheme.

Figure 6.2 shows the components of $\hat{D}$ with error bars of length $2\widehat{\sigma}_{\hat{D}}$. This range typically covers at least 75% of a distribution, and usually more (95% for a normal distribution). We also show the same analysis applied to the minimum system cost (the *min* of the problem, instead of the *argmin*). The $\hat{D} \pm 2\widehat{\sigma}_{\hat{D}}$ ranges are an imperfect but simple way of visualising uncertainty. They do not convey all information about possible values. For example, they may extend to negative values even when capacities and costs must be positive, and ignore any *covariance*

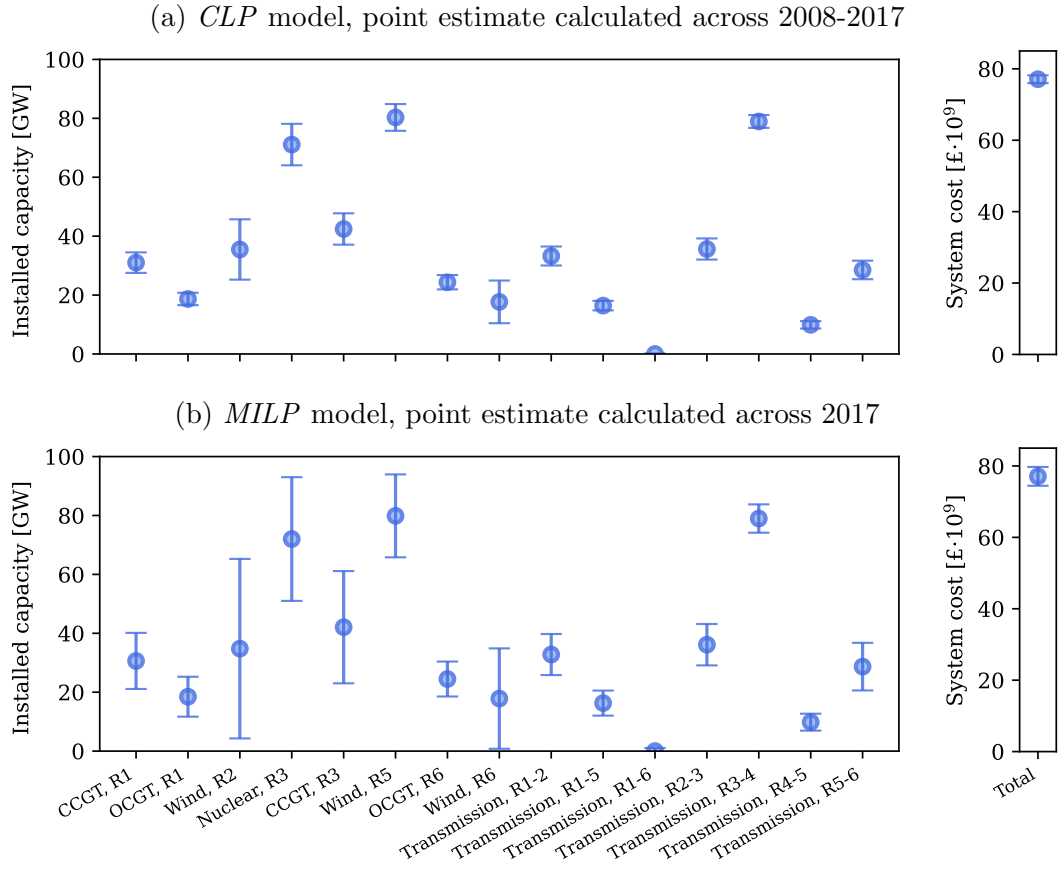


Figure 6.2: Point estimates $\hat{D}$ of selected model outputs, with error bars $\pm 2\hat{\sigma}_{\hat{D}}$ (2 standard deviations). $\hat{\sigma}_{\hat{D}}$ is estimated using the m -out-of- n bootstrap with $K=1000$ bootstrap samples. See Table [6.1](#) for details.

structures. However, they provide a simple visualisation of relative variabilities for a large range of output types.

For the *CLP* model in Figure 6.2(a), the use of a ten-year sample for the point estimate means uncertainty levels are small. Cost-optimal wind capacity in Region 2 has a $\hat{D} \pm 2\widehat{\sigma}_D$ range of 25-45GW, but other outputs, such as other generation and transmission capacities and the minimum system cost, are comparatively certain.

For the *MILP* model in Figure 6.2(b), the use of a shorter (one-year) point estimate means uncertainty bounds are large. For example, $\hat{D} \pm 2\widehat{\sigma}_D$ extends from 2-62GW for optimal wind capacity in Region 2. We cannot say with any certainty whether cost-optimal system design has more CCGT capacity in Regions 1 or 3, as the uncertainty regions overlap. In test case (a), in contrast, we have a high confidence that more CCGT capacity should be installed in Region 3. In both (a) and (b), optimal transmission capacity and minimum system cost are comparatively certain.

The *m-out-of-n* bootstrap is more data-efficient than disjoint Monte Carlo (using disjoint samples of the same length as the point estimate). For example, in (b), disjoint Monte Carlo across $K=1000$ samples requires 1000 years of demand and weather data. In (a), it requires 10000 ($=10 \times 1000$) years. Instead, we resample from the ten years we have, requiring no additional data.

The *m-out-of-n* bootstrap also significantly reduces the computing requirements. In test case (a), the proposed method represents a 44-fold reduction in computing time (six vs. 262 minutes) and an eight-fold reduction in memory (1633 vs. 13813 MB). Since the estimates are generated across $K=1000$ samples, the total computing time in this case is 6000 minutes (100 hours), compared with 262000 (≈ 182 days) using standard Monte Carlo methods. The effect is even more pronounced in test case (b), where the method amounts to a roughly 55-fold (14 vs. 765 minute) reduction.

6.4.3 Determining required sample length

The relationship between sample length and standard deviation allows an informed choice of sample length based on desired certainty levels. For example, suppose we want $\widehat{\sigma}_D$ on optimal nuclear capacity in Region 3 to be no more than 5GW. In test case (a), the method indicates that σ_S , the standard deviation across one-year bootstrap samples, is about 11GW. Using equation (6.4) then indicates a required sample length of $n = 1 \times \frac{11^2}{5^2} \approx 5$ years for the point estimate.

6.4.4 Verifying the method's consistency

In Sections 6.4.2 and 6.4.3, we construct standard deviation estimates under the assumption that they are consistent. This section examines whether this is actually the case. We use the m -out-of- n bootstrap, with varying subsample length m , to estimate σ_Y , the sampling standard deviation of design estimates calculated across single years. We also construct a 95% confidence interval for σ_Y using the 38 individual (disjoint) years from 1980-2017 and the bootstrap *percentile* method, a well-established procedure [Van der Vaart, 2014, Chapter 23]. If the method is consistent, the standard deviation estimates (1) lie within the confidence interval with high probability, and (2) are independent of m , indicating that the $\frac{m}{n}$ scaling factor used in Step 4 of Section 6.3.1 is correct. These conditions are necessary but not sufficient for the method's consistency; however, we will see that they provide a useful heuristic.

Figure 6.3 shows the estimates $\hat{\sigma}_Y$ along with a 95% confidence interval for the true value σ_Y . Estimates using subsamples shorter than one year ($m=4, 12, 24, 36$ and 48 weeks) are generated using the *weeks* scheme, while longer subsamples ($m=1, 2$ and 3 years) are generated via the *months* scheme. For the *MILP* model (test case (b)), the longest sample length is one year since the integer constraints make longer samples computationally unfeasible.

We consider the design estimates (installed capacities of generation and transmission technologies) as well as a number of functions of them: the minimum system cost, carbon emissions, as well as the peak unmet demand: the maximum supply shortfall across the time series. In Figure 6.3, we show aggregated capacities (sums across the regions) to reduce the number of plots.

Figure 6.3 indicates that, with sufficient subsample size, the standard deviation estimates for model outputs meet the two required conditions; they lie within the 95% confidence interval and are roughly independent of sample length (as seen by a constant y -value). The one output where estimates are clearly not consistent is the peak unmet demand. The reason for this is, as discussed in Section 6.3.4, that this model output is a sample maximum, namely that of demand minus available generation. Indeed, we included this output specifically to highlight where the m -out-of- n bootstrap method is *not* consistent.

The approximate diagnostic introduced in Section 6.3.4 — that $\hat{\sigma}_Y$ is independent of subsample size m — successfully identifies when standard deviation estimates are consistent. This diagnostic implies that results are approximately valid for all

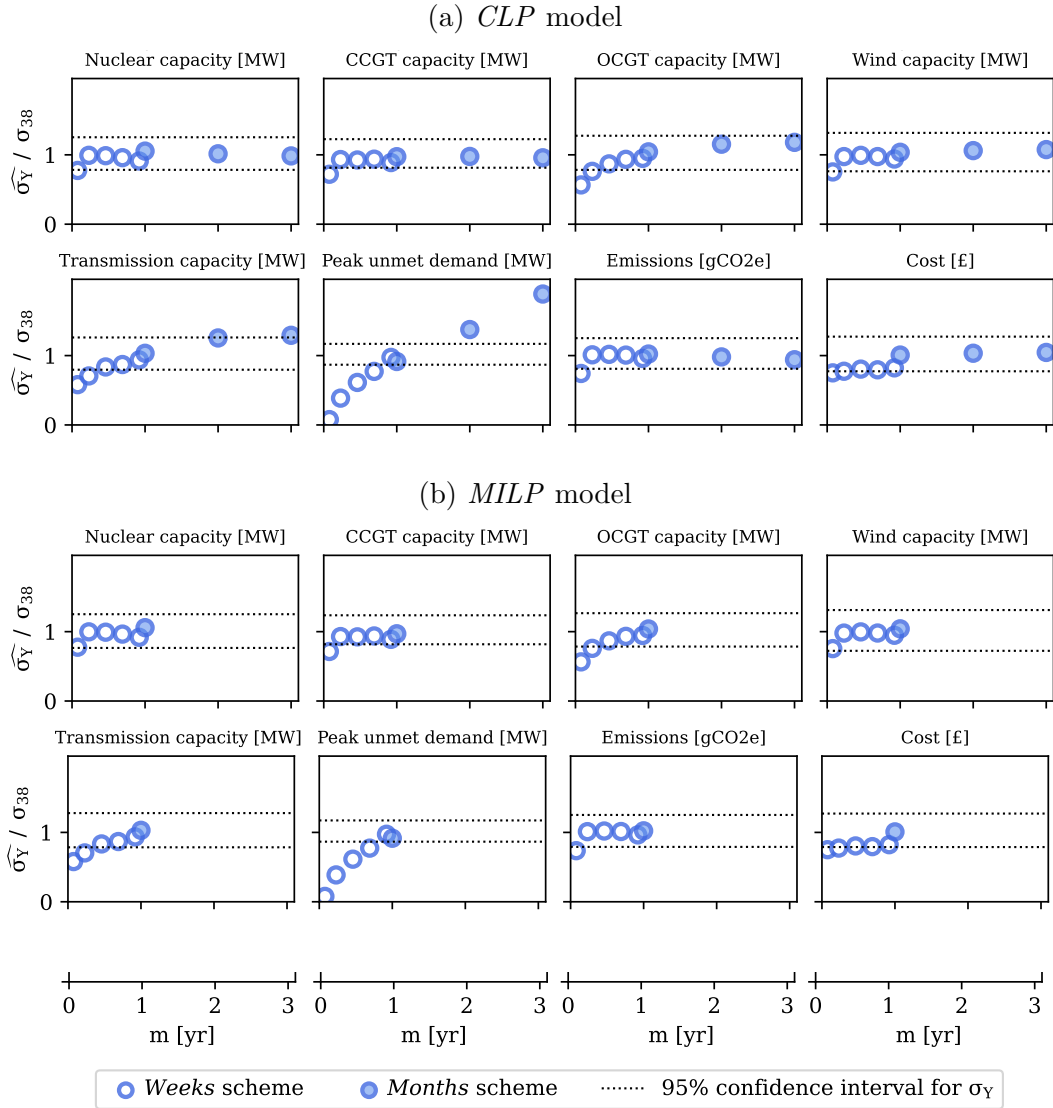


Figure 6.3: Single-year standard deviation estimates $\widehat{\sigma}_Y$, calculated across $K=1000$ bootstrap samples, as a function of subsample size m . Dashed lines: symmetric 95% confidence interval for true value σ_Y . σ_{38} : standard deviation across the 38 individual years from 1980-2017. All plots share the same x -axis.

outputs except peak unmet demand. These are precisely the model outputs that consistently lie within the confidence interval.

Two further observations warrant mention. The first is that the method fails when subsample sizes are too small, as for transmission capacity. The second is that the *weeks* scheme usually generates slightly lower standard deviation estimates than the *months* scheme and may underestimate the target σ_Y . This most likely comes from small levels of dependence at time scales longer than seven days (see Figure 6.1) which is treated as zero when resampling blocks, artificially reducing the standard deviation (sometimes called *whitening*). The associated error rarely exceeds 15%. Furthermore, the independent resampling of blocks from different years does not seem to lead to much underestimation of variance by smoothing over inter-year variability as described by Bodini et al. 2016.

6.4.5 Discussion

The simulation study highlights a number of messages. The first is that the method usually provides consistent uncertainty estimates on model outputs at reduced computational cost and without requiring any additional data. It fails when the subsample size is too small (as is virtually always the case in computational uncertainty quantification) or when the model output depends strongly on a sample maximum (e.g. maximum or peak unmet demand). Each of these failures are identified by the diagnostic introduced in Section 6.3.4, giving users an indication whether estimates can be trusted. Furthermore, the method provides consistent estimates for nuclear capacity for the *MILP* model, even though this is a discrete variable in blocks of 3GW — probably because the jumps are small enough to be well approximated by a continuous variable.

The case study illustrates, in accordance with previous studies (Section 2.5), the risk in using short demand and weather samples in power system planning models. For example, the cost optimal system design for the *CLP* model contains 15GW of wind capacity in Region 2 under 1982 data, but 92GW under 1986 data, a more than sixfold difference. For robust planning outputs under demand and weather uncertainty, samples spanning multiple years should be considered.

We make three recommendations for users of this method. The first is to make subsamples used for the point estimate as long as possible, as this reduces the uncertainty on their values. The second is that, if both are computationally feasible, the *months* scheme should be preferred over the *weeks* scheme, as the former has

less tendency to underestimate variability. The third is to estimate the standard deviation with multiple subsample lengths and check if uncertainty estimates are close, since a failure for this diagnostic indicates that estimates are inconsistent.

6.5 Conclusions

In this chapter, we introduce a scheme to efficiently quantify the impact of forward propagated demand and weather uncertainty in power system models. It provides an information advantage over point estimates by indicating whether an output is statistically robust or a result of the particular demand and weather sample. The method is applicable to planning models with simulations of at least one year and without constraints (e.g. storage) that links the sampled blocks (weeks or months). Simulation lengths may be informed based on desired uncertainty levels, avoiding needless computational expense. The method can be expected to generate consistent uncertainty estimates for most PSM outputs, and an approximate diagnostic can help determine (but not prove) whether this is the case.

We recommend a number of extensions. One is to combine our method with uncertainty analysis on other model outputs (e.g. technology or fuel costs) by sampling both demand and weather data and other uncertain input parameters simultaneously. Another extension is the determination of confidence intervals or the *covariance* of model outputs instead of the standard deviation for each model output individually. Adapting the method for models with seasonal storage is also a natural generalisation. Finally, the method could leverage the use of subsampling methods for point estimates as in Chapter 4. Such approaches can be straightforwardly combined with the algorithm proposed in this paper by using subsampled data for the point estimate. However, more sophisticated methods, such as using subsampled data in the bootstrap samples, or accounting for the uncertainty induced by subsampling for the point estimate, require additional modifications.

7. Conclusions

This chapter concludes the thesis. We review research context, our contributions and their implications. We also recommend a number of extensions.

7.1 Context

Decarbonisation efforts make power (electricity) systems increasingly weather-dependent. In particular, variable renewables such as solar and wind, whose generation cannot be controlled, represent an increasing share of supply. This complicates the planning and operation of power systems, which require a continuous supply and demand balance.

In this thesis, we study *power system planning models* that inform investments into generation, transmission and storage infrastructure. These models estimate optimal system design by minimising a cost function given time series data such as demand levels, wind speeds and solar irradiances. Model outputs are subject to uncertainty, since future demand and weather patterns are unknown and will differ from those historically observed.

The high computational cost of solving realistic power system planning models limits their effectiveness. This is usually possible only across relatively short simulation lengths (a few days or weeks), generating model outputs that are poor estimates for their true values (Section 2.5). Furthermore, the analysis of future climate on power systems — typically using multiple decades of data from a climate model — becomes impossible.

Many previous approaches to reduce computational cost in planning models induce significant errors. These typically originate from a reduction in variability and extreme events (Section 4.1.4).

7.2 Contributions

In this thesis, we introduce algorithms to mitigate, in a computationally efficient manner, the impact of demand and weather uncertainty in power system planning problems. Our methods solve planning models across subsamples of long time series, carefully chosen to induce less approximation error than established approaches. Our contributions can be grouped into two classes, described below.

The first class of contributions makes optimal design estimates closer to the long-run optima without sacrificing computational feasibility. The key innovation is a systematic identification and preservation of relevant extreme events. We introduce this algorithm in Chapter 4 and extend it to models with storage in Chapter 5. Mathematically, the method is an adaptive importance sampling scheme customised to power system planning models.

The second class of contributions efficiently quantifies the sampling uncertainty on design estimates (Chapter 6), giving an indication whether results are statistically robust or an artefact of the particular demand and weather sample. Mathematically, our method is an m -out-of- n time series block bootstrap that acts on *shorter* time series and adjusts uncertainty bounds accordingly. This allows Monte Carlo analysis across runs with lower computational cost.

We also release a set of open-source power system planning models designed for the analysis of demand and weather impacts. These models, along with time series data, processing code and a tutorial, are available at doi:[10.5281/zenodo.7025098](https://doi.org/10.5281/zenodo.7025098).

This thesis combines and extends techniques from three distinct research areas: power system modelling, optimisation and statistics. From power system modelling, we use open-source software and solvers and build upon established *time series aggregation* (Sections 4.1.1 and 5.1.1) approaches. We use solution algorithms from the optimisation literature. Finally, we build upon a number of statistical techniques, including Monte Carlo methods, importance sampling and the bootstrap. The interpretation of planning models as stochastic programming problems (Section 2.7) provides the conceptual link.

7.3 Implications

Our methods allow the consideration of longer samples in power system planning models than was hitherto possible. Previous time series aggregation methods (Sections 4.1.1 and 5.1.1) do the same, but induce significant errors in model outputs. Furthermore, we introduce uncertainty quantification methods that indicate how confident users can be in their modelling results.

Improvements to the accuracy of, and understanding of the uncertainties around, planning model outputs is important. Previous research (Section 2.5) indicates that key conclusions — such as whether sustainable energy requires nuclear, biomass or carbon capture and storage to remain secure and affordable — can change depending on the choice of demand and weather data. Our methods make answers to such

questions more robust.

7.4 Extensions

We see a number of natural extensions to this thesis. Perhaps the most obvious is the generalisation of the uncertainty quantification scheme in Chapter 6 to models with storage, as we do for the point estimates between Chapters 4 and 5. This is a nontrivial task; our current approach relies on using *shorter* time series, while storage level aggregation requires chronology across the *full* sample. One possible approach is to adjust time step weights, e.g. double changes in storage levels when two temporal periods are reduced to one.

Another generalisation is the combination of methods for improving point estimates (Chapters 4 and 5) with that for uncertainty quantification (Chapter 6). In the uncertainty quantification scheme, we perform Monte Carlo analysis across shorter samples than the original, created by block bootstrapping. If we subsequently aggregate these samples into even shorter periods using the methods from Chapters 4 and 5, we may be able to consider longer bootstrap resamples, enhancing standard deviation estimates.

In all our case studies, we use 38 years of historic demand and weather data. For this reason, all the events we consider occur in this dataset — itself a sample from some underlying distribution. A valuable extension is the consideration of out-of-sample events or the use a far larger base dataset. One way to do this is to create artificial data using a climate model, which also allows us to explore the impacts of e.g. very rare events or climate change.

A more general extension is to examine uncertainties not originating from demand and weather time series inputs. Some examples, such as generation or transmission failures, admit similar statistical analysis. For others, such as future technological or price developments, electrification rates of new industries (e.g. transport and heating) or changes in human behaviour, a statistical model is less obvious and scenario-driven analysis may be required.

Finally, we can develop mathematical theory for our methods. This may involve precise conditions for our methods to work, as well as e.g. convergence rates or error bounds.

A. Additional scientific background

In this appendix, we expand the scientific background (Section 2). We discuss Monte Carlo methods (Appendix A.1), importance sampling (Appendix A.2), bootstrap methods (Appendix A.3) and statistical theory for the *sample average approximation* (Appendix A.4).

A.1 Monte Carlo methods

Monte Carlo methods use repeated sampling to estimate quantities that depend on a probability distribution [Metropolis and Ulam, 1949]. Let $\boldsymbol{\xi} \in \Xi$ be a random variable with distribution $\mathbb{P}$, and let ϕ^* be a derived quantity, such as the probability of an event or the expectation of a random variable.

Monte Carlo methods are useful when we cannot calculate ϕ^* directly (because the calculation is complex, or we do not know the underlying distribution) but we can *sample* from $\mathbb{P}$. We estimate ϕ^* by $\hat{\phi}_n(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n)$, where $\hat{\phi}_n : \Xi^n \rightarrow \mathbb{R}$ is a statistical *estimator* (real-valued function) and $\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n$ is a sample drawn from $\mathbb{P}$. For now, we assume the $\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n$ are independent and identically distributed (IID) and use the notation

$$\hat{\phi}_n(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n) \quad | \quad \boldsymbol{\xi}_i \sim \mathbb{P}. \quad (\text{A.1})$$

to denote $\hat{\phi}_n(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n)$ with samples $\boldsymbol{\xi}_i$ drawn IID from $\mathbb{P}$.

As an example, suppose we estimate the unknown probability p that a (biased) coin lands heads up. Let $\boldsymbol{\xi}$ be a random variable with value 1 for heads (probability p) and 0 for tails (probability $1 - p$). In the formalism above, a natural estimate for $\phi^* = p = \mathbb{E}(\boldsymbol{\xi})$ is the rate of heads across n coin tosses:

$$\hat{\phi}_n(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n) = \frac{1}{n} \sum_{i=1}^n \xi_i \quad (\text{A.2})$$

where the ξ_i are results of the n coin tosses. This is the *sample mean* or *sample average* estimator. In the next section, we discuss a theoretical result that validates this approach.

A.1.1 Expected values

Estimating expected values, as we do in the biased coin problem above, is a common application of Monte Carlo methods. Let $F : \Xi \rightarrow \mathbb{R}$ be a real-valued function of ξ . It has mean

$$\mu_F := \mathbb{E}(F(\xi)) = \int_{\Xi} F(\xi) d\mathbb{P}(\xi) \quad (\text{A.3})$$

where $d\mathbb{P} = d\mathbb{P}(\xi)$ is the probability measure evaluated at ξ (we abuse notation by using ξ as both a random variable and a particular value of it, as an integration variable). In the coin toss example, $F(\xi) \equiv \xi$ and $\mu_F = p$.

Given a sample $\xi_1, \dots, \xi_n$ drawn IID from $\mathbb{P}$, a natural estimate comes from the *sample mean* (or *sample average*):

$$\hat{\phi}_n(\xi_1, \dots, \xi_n) = \frac{1}{n} \sum_{i=1}^n F(\xi_i). \quad (\text{A.4})$$

Its mean and variance are

$$\mathbb{E}(\hat{\phi}_n) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(F(\xi_i)) = \mu_F \quad (\text{A.5})$$

$$\text{Var}(\hat{\phi}_n) = \frac{1}{n^2} \sum_{i=1}^n \mathbb{E}((F(\xi_i) - \mu_F)^2) = \frac{\sigma_F^2}{n} \quad (\text{A.6})$$

where $\sigma_F^2 := \text{Var}(F) := \mathbb{E}((F(\xi) - \mu_F)^2)$. Hence, $\hat{\phi}_n$ is an unbiased estimator for μ_F with a variance decreasing linearly with sample size n . For this reason, we might expect $\hat{\phi}_n$ to approximate $\phi^* = \mu_F$ with increasing accuracy as $n \rightarrow \infty$.

A number of theoretical results support this approach. The *weak law of large numbers* states that if $\sigma_F^2 < \infty$ (finite variance), then $\hat{\phi}_n$ converges to μ_F in probability as $n \rightarrow \infty$. The *central limit theorem* provides additional information for finite values of n . Consider the scaled and shifted quantity

$$T_n(\hat{\phi}_n, \mu_F) := \sqrt{n}(\hat{\phi}_n(\xi_1, \dots, \xi_n) - \mu_F). \quad (\text{A.7})$$

The central limit theorem states that, if $\sigma_F^2 < \infty$ (finite variance),

$$\sqrt{n}(\hat{\phi}_n - \mu_F) \xrightarrow{d} \mathcal{N}(0, \sigma_F^2) \quad \text{as } n \rightarrow \infty. \quad (\text{A.8})$$

where $\mathcal{N}(0, \sigma_F^2)$ is a normal distribution with mean 0 and variance σ_F^2 . The conver-

gence is in *distribution* as opposed to the stronger convergence in *probability* for the weak law of large numbers.

The asymptotic distribution of T_n provides approximate bounds on μ_F . Let

$$\hat{G}_n(\lambda) := \mathbb{P}\left(\sqrt{n}(\hat{\phi}_n(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n) - \mu_F) \leq \lambda\right) \quad (\text{A.9})$$

be the cumulative distribution function (CDF) of T_n . From the central limit theorem, for large n ,

$$\hat{G}_n(\lambda) \approx \Phi\left(\frac{\lambda}{\sigma_F}\right) = \int_{-\infty}^{\frac{\lambda}{\sigma_F}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \quad (\text{A.10})$$

which is the CDF of the standard normal $\mathcal{N}(0, 1)$, whose values can be approximated reliably. Equation (A.9) then gives a confidence interval for μ_F :

$$\mathbb{P}(\mu_F \in [\eta_1, \eta_2]) \approx \Phi\left(\frac{\sqrt{n}}{\sigma_F}(\eta_2 - \hat{\phi}_n)\right) - \Phi\left(\frac{\sqrt{n}}{\sigma_F}(\eta_1 - \hat{\phi}_n)\right) \quad (\text{A.11})$$

for all $\eta_1 < \eta_2$. Given a sample $\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n$, we can calculate $\hat{\phi}_n$ and use (A.11) to construct approximate confidence intervals for μ_F . The value σ_F is usually estimated from the data. A t -distribution, which considers approximation error of the sample standard deviation, may also be appropriate.

A.1.2 Other use cases

In the previous section we explain how confidence bounds on the population mean can be created via the sample mean. However, Monte Carlo methods are also useful in many other use cases. One relevant example is the consideration of estimators $\hat{\phi}_n(\mathbf{x}, \boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n)$ that are a function of an additional argument $\mathbf{x}$. In such cases, we seek results that hold for all $\mathbf{x}$ in a domain $\mathcal{X}$. Another example consists of M -estimators: minimisers of sample average objective functions. These two classes of estimators are closely linked to sample average approximations of stochastic programming problems as in Section 2.6: the objective function is an estimator that depends on both a sample $\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n$ and a parameter $\mathbf{x}$, while the associated minimiser is an M -estimator. We discuss this link in more detail in Appendix A.4.

Monte Carlo methods may also be used for uncertainty analysis, where we seek not a population parameter, but the distribution of the estimator itself across samples. The theoretical justification for this approach, in which the estimator is calculated across different samples to infer its distribution, is the *Glivenko-Cantelli*

theorem and its various extensions, which demonstrate that (under certain conditions) estimators' empirical distribution functions converge to their probability distributions. [Van der Vaart 2014, Chapter 19] provides a detailed discussion.

A.2 Importance sampling

Importance sampling is a Monte Carlo method that samples from a different distribution than $\mathbb{P}$ [Tokdar and Kass, 2010]. This is usually done to reduce an estimator's variance. Let $\mathbb{Q}$ be a probability distribution that satisfies $d\mathbb{Q}(\boldsymbol{\xi}) > 0$ whenever $d\mathbb{P}(\boldsymbol{\xi}) > 0$. Consider the *importance sampling estimator* for the sample mean:

$$\hat{\phi}_n^{\text{IS}}(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n) = \frac{1}{n} \sum_{i=1}^n F(\boldsymbol{\xi}_i) \frac{d\mathbb{P}(\boldsymbol{\xi}_i)}{d\mathbb{Q}(\boldsymbol{\xi}_i)} \quad | \quad \boldsymbol{\xi}_i \sim \mathbb{Q} \quad (\text{A.12})$$

where the $\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n$ are drawn IID from $\mathbb{Q}$. The factor $\frac{d\mathbb{P}(\boldsymbol{\xi})}{d\mathbb{Q}(\boldsymbol{\xi})}$ is the *likelihood ratio*. By construction, the expectation of $\hat{\phi}_n^{\text{IS}}$ under sampling from $\mathbb{Q}$ is μ_F :

$$\mathbb{E}(\hat{\phi}_n^{\text{IS}}) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}_{\mathbb{Q}} \left(F(\boldsymbol{\xi}_i) \frac{d\mathbb{P}(\boldsymbol{\xi}_i)}{d\mathbb{Q}(\boldsymbol{\xi}_i)} \right) = \int_{\Xi} F(\boldsymbol{\xi}_i) \frac{d\mathbb{P}(\boldsymbol{\xi}_i)}{d\mathbb{Q}(\boldsymbol{\xi}_i)} d\mathbb{Q}(\boldsymbol{\xi}_i) = \mu_F. \quad (\text{A.13})$$

The variance is

$$\text{Var}(\hat{\phi}_n^{\text{IS}}) = \frac{1}{n^2} \sum_{i=1}^n \mathbb{E}_{\mathbb{Q}} \left(\left(F(\boldsymbol{\xi}_i) \frac{d\mathbb{P}(\boldsymbol{\xi}_i)}{d\mathbb{Q}(\boldsymbol{\xi}_i)} - \mu_F \right)^2 \right) \quad (\text{A.14})$$

$$= \frac{1}{n} \int_{\Xi} \left(F(\boldsymbol{\xi}_i) \frac{d\mathbb{P}(\boldsymbol{\xi}_i)}{d\mathbb{Q}(\boldsymbol{\xi}_i)} - \mu_F \right)^2 d\mathbb{Q}(\boldsymbol{\xi}_i) \quad (\text{A.15})$$

$$= \frac{1}{n} \left[\int_{\Xi} F(\boldsymbol{\xi}_i)^2 \frac{d\mathbb{P}(\boldsymbol{\xi}_i)}{d\mathbb{Q}(\boldsymbol{\xi}_i)} d\mathbb{P}(\boldsymbol{\xi}_i) - \mu_F^2 \right]. \quad (\text{A.16})$$

This suggests a well-chosen *importance distribution* $\mathbb{Q}$ can reduce the variance of $\hat{\phi}_n^{\text{IS}}$. In particular, the *zero-variance density*

$$d\mathbb{Q}(\boldsymbol{\xi}) = \frac{F(\boldsymbol{\xi}) d\mathbb{P}(\boldsymbol{\xi})}{\int_{\Xi} F(\boldsymbol{\xi}') d\mathbb{P}(\boldsymbol{\xi}')} \quad (\text{A.17})$$

is valid when $F(\boldsymbol{\xi}) \geq 0$ everywhere and gives $\hat{\phi}_n^{\text{IS}}$ a variance of 0. It is typically unusable since it requires calculation of the normalising constant $\int_{\Xi} F(\boldsymbol{\xi}') d\mathbb{P}(\boldsymbol{\xi}')$, often the quantity being estimated in the first place. However, it suggests samples should be weighted roughly by their contribution to the integral and their contribution subsequently reduced via the likelihood ratio.

Importance sampling methods are useful when $\hat{\phi}_n$ has high variance for small n but is costly to evaluate for large n . In this case, it can reduce variance without an increase in sample size. Intuitively, it works by including high impact, low probability events more often and correcting for their oversampling through the likelihood ratio. To do this, we need the ability to sample from $\mathbb{Q}$ and the value of $\frac{d\mathbb{P}(\boldsymbol{\xi})}{d\mathbb{Q}(\boldsymbol{\xi})}$ at each sample value.

The above discussion considers the sample mean estimator. However, importance sampling may also be applied to other estimators. For example, Glynn [1996] create importance sampling estimators for empirical quantiles.

A.3 Bootstrap methods

Bootstrap methods Efron [1979] use resampling for statistical inference. In its most basic form, they emulate an estimator's sampling distribution by its distribution under resampling from a fixed data set. As in Appendix A.1, let $\hat{\phi}_n(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n)$ be an estimator for ϕ^* with $\boldsymbol{\xi}_i$ drawn IID from distribution $\mathbb{P}$:

$$\hat{\phi}_n(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n) \quad | \quad \boldsymbol{\xi}_i \sim \mathbb{P}. \quad (\text{A.18})$$

We can deduce properties of ϕ^* from an estimator $T_n(\hat{\phi}_n, \phi^*)$ with a known (non-degenerate) limiting distribution as $n \rightarrow \infty$. For example, for the sample mean in Appendix A.1.1, $T_n(\hat{\phi}_n, \phi^*) = \sqrt{n}(\hat{\phi}_n - \phi^*) \xrightarrow{d} \mathcal{N}(0, \sigma_F^2)$ (convergence in distribution), which we use to construct confidence intervals for $\phi^* = \mu_F$. Similar results, with different limit theorems, apply to other estimators, such as empirical quantiles Van der Vaart [2014, chapter 21].

Bootstrap methods give limiting distributions for $T_n(\hat{\phi}_n, \phi^*)$ for a large class of estimators $\hat{\phi}_n$ without having to use a separate limit theorem in each case. They use the *bootstrap estimator*

$$\hat{\phi}_n(\boldsymbol{\xi}_1^B, \dots, \boldsymbol{\xi}_n^B) \quad | \quad \boldsymbol{\xi}_i^B \sim \hat{\mathbb{P}}_n^B(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n). \quad (\text{A.19})$$

where $\hat{\mathbb{P}}_n^B(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n)$ is the *bootstrap distribution*. In the simplest case, $\hat{\mathbb{P}}_n^B$ is the *empirical distribution* of $\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n$, and the $\boldsymbol{\xi}_i^B$ are sampled with replacement from $\{\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n\}$. The bootstrap estimator has two sources of randomness: one for the sample $\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n \sim \mathbb{P}$ and a second for the bootstrap resample $\boldsymbol{\xi}_1^B, \dots, \boldsymbol{\xi}_n^B \sim \hat{\mathbb{P}}_n^B$. When considering (A.19), we usually fix $\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n$ and consider the distribution under *resampling* from $\hat{\mathbb{P}}_n^B$.

A.3.1 Machinery

The driving result of bootstrap methods is that, under certain conditions, the following estimators have the same limiting distribution as $n \rightarrow \infty$:

$$T_n(\hat{\phi}_n(\xi_1, \dots, \xi_n), \phi^*) \mid \xi_i \sim \mathbb{P} \quad (\text{A.20})$$

$$T_n(\hat{\phi}_n(\xi_1^B, \dots, \xi_n^B), \hat{\phi}_n(\xi_1, \dots, \xi_n)) \mid \xi_i^B \sim \hat{\mathbb{P}}_n^B. \quad (\text{A.21})$$

In (A.20), the ξ_i are sampled from $\mathbb{P}$. In (A.21), the ξ_i^B are sampled from $\hat{\mathbb{P}}_n^B = \hat{\mathbb{P}}_n^B(\xi_1, \dots, \xi_n)$ with $\xi_1, \dots, \xi_n$ fixed. Since every sample $\xi_1, \dots, \xi_n$ determines its own distribution (A.21), the definition of bootstrap *consistency* is usually the distribution of (A.21) converging to that of (A.20) with probability 1 across samples $\xi_1, \dots, \xi_n \sim \mathbb{P}$ as $n \rightarrow \infty$ [Van der Vaart, 2014, chapter 23].

Bootstrap methods *plug in* the bootstrap version of T_n (A.21) for the original (A.20). Recall that the distribution function of $T_n(\hat{\phi}_n, \phi^*)$ gives us properties of ϕ^* (e.g. a confidence interval for its value) and $\hat{\phi}_n$ (e.g. its sampling standard deviation). We estimate it using the bootstrap as follows:

$$\mathbb{P} \left(T_n \left(\hat{\phi}_n(\xi_1, \dots, \xi_n), \phi^* \right) \leq \lambda \right) \quad (\text{A.22})$$

$$\approx \hat{\mathbb{P}}_n^B \left(T_n \left(\hat{\phi}_n(\xi_1^B, \dots, \xi_n^B), \hat{\phi}_n(\xi_1, \dots, \xi_n) \right) \leq \lambda \right) \quad (\text{A.23})$$

$$= \frac{1}{n^n} \sum_{i_1=1}^n \cdots \sum_{i_n=1}^n \mathbb{1} \left(T_n \left(\hat{\phi}_n(\xi_{i_1}, \dots, \xi_{i_n}), \hat{\phi}_n(\xi_1, \dots, \xi_n) \right) \leq \lambda \right). \quad (\text{A.24})$$

where $\mathbb{1}(\ast)$ takes value 1 when $\ast$ is true and 0 otherwise. From the first to the second line, we use the bootstrap's consistency, i.e. that (A.20) and (A.21) have the same asymptotic distribution. In the last line, we use that the bootstrap distribution is resampling with replacement from $\{\xi_1, \dots, \xi_n\}$. Since n^n increases very quickly, we usually use a Monte Carlo approximation with a random selection of bootstrap resamples.

Bootstrap consistency depends on properties of the estimator $\hat{\phi}_n$ and distribution $\mathbb{P}$. [Singh, 1981] derives weak conditions for bootstrap consistency in sample means and quantiles. [Bickel and Freedman, 1981] provide similar results for other estimators and discuss general conditions and counterexamples. [Davison and Hinkley, 1997] describes refinements for time series data, when samples are not independent and identically distributed.

A.3.2 The m -out-of- n bootstrap

The bootstrap estimator is usually calculated across resamples of the same length as the original dataset, i.e. using

$$T_n(\hat{\phi}_n^B(\boldsymbol{\xi}_1^B, \dots, \boldsymbol{\xi}_n^B), \hat{\phi}_n) \mid \boldsymbol{\xi}_i^B \sim \hat{\mathbb{P}}_n^B(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n). \quad (\text{A.25})$$

When calculating $\hat{\phi}_n^B(\boldsymbol{\xi}_1^B, \dots, \boldsymbol{\xi}_n^B)$ is expensive, we may also use a smaller number of samples $m < n$. This implies the bootstrap estimator

$$T_m(\hat{\phi}_m^B(\boldsymbol{\xi}_1^B, \dots, \boldsymbol{\xi}_m^B), \hat{\phi}_n) \mid \boldsymbol{\xi}_i^B \sim \hat{\mathbb{P}}_n^B(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n). \quad (\text{A.26})$$

This is the m -out-of- n bootstrap. Instead of $\hat{\phi}_n^B$, we calculate $\hat{\phi}_m^B$ repeatedly, which is usually cheaper to evaluate. (A.26) has the same limiting distribution as (A.25) whenever the traditional (n -out-of- n) bootstrap is consistent, and in some cases when it is not. Bickel et al. [1997] discuss theoretical properties.

For many such estimators, $T_n(\hat{\phi}_n, \phi^*)$ is of the form $\sqrt{n}(\hat{\phi}_n - \phi^*)$ and the m -out-of- n bootstrap estimator becomes

$$T_m(\hat{\phi}_m^B, \hat{\phi}_n) = \sqrt{m}(\hat{\phi}_m^B - \hat{\phi}_n). \quad (\text{A.27})$$

In this case, we may use $\hat{\phi}_m^B$ to estimate prediction intervals or the sampling standard deviation for $\hat{\phi}_n$ provided we rescale by the factor $\sqrt{\frac{n}{m}}$ to correct for the reduction in sample length. This is what we do in Chapter 6.

A.4 Sample average approximation: theory

In this section, we briefly introduce statistical theory for the *sample average approximation* to stochastic programming problems (Section 2.6.1). Let θ^* and $\mathbf{x}^*$ be the minimum and minimiser of a stochastic programming problem, and let $\hat{\theta}_n = \hat{\theta}_n(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n)$ and $\hat{\mathbf{x}}_n = \hat{\mathbf{x}}_n(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n)$ be the corresponding quantities for the sample average approximation. We view $\hat{\theta}_n$ and $\hat{\mathbf{x}}_n$ as statistical estimators for θ^* and $\mathbf{x}^*$ and investigate their properties below. See Homem-de Mello and Bayraksan, 2014 for additional details.

Much statistical theory for sample average approximations are inherited from M -estimators. For this reason, we introduce M -estimators in Section A.4.1, before examining theoretical results for sample average approximations in Section A.4.2.

A.4.1 M-estimators

An *M-estimator* [Huber, 1964], minimises the sample mean of a function $F(\mathbf{x}, \boldsymbol{\xi})$ that depends on both a parameter $\mathbf{x} \in \mathbb{R}^{d_{\mathbf{x}}}$ and a random variable $\boldsymbol{\xi} \in \Xi$ [Van der Vaart, 2014]. Let $\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n \sim \mathbb{P}$ be a sample from Ξ^n , and define

$$\hat{f}_n(\mathbf{x} \mid \boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n) = \frac{1}{n} \sum_{i=1}^n F(\mathbf{x}, \boldsymbol{\xi}_i) \quad (\text{A.28})$$

to be the sample mean of F . An *M-estimator* is a function $\hat{\mathbf{x}}_n = \hat{\mathbf{x}}_n(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n)$ that minimises (A.28) under the first argument $\mathbf{x}$.

M-estimators generalise various statistical functions. If $\mathbf{x} = x$ and $\boldsymbol{\xi} = \xi$ are one-dimensional, the sample mean and median fit this framework with $F = (x - \xi)^2$ and $|x - \xi|$ respectively. If $\boldsymbol{\xi} = (\mathbf{X}, y)$ contains matched independent ($\mathbf{X}$) and dependent (y) variables, then ordinary least squares regression has $F = (y - a - \mathbf{b}^T \mathbf{X})^2$, where a and $\mathbf{b}$ are the intercept and coefficients respectively.

In fact, *M-estimators* are generalisations of any *maximum likelihood estimator* (MLE) of IID samples. Let $p(\boldsymbol{\xi}|\mathbf{x})$ be the probability density function (PDF) of the random variable $\boldsymbol{\xi}$ given the parameter $\mathbf{x}$. MLEs minimise the (mean) *negative loglikelihood*

$$l(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n | \mathbf{x}) := -\frac{1}{n} \sum_{i=1}^n \log p(\boldsymbol{\xi}_i | \mathbf{x}) \quad (\text{A.29})$$

which corresponds to (A.29) with $F(\mathbf{x}, \boldsymbol{\xi}) = -\log p(\boldsymbol{\xi}|\mathbf{x})$. [Van der Vaart, 2014, Chapter 5] provides a more detailed discussion.

A.4.2 Sample average approximations

Since the minimisers of sample average approximations are *M-estimators*, their theory is closely linked. Under a set of assumptions on F and $\mathbb{P}$, $\hat{\theta}_n$ and $\hat{\mathbf{x}}_n$ are consistent for θ^* and $\mathbf{x}^*$ respectively. [Wald, 1949] and [Van der Vaart, 2014, Chapter 5.2] provide simple proofs of the consistency of maximum likelihood estimators and *M-estimators* respectively. [Shapiro, 1989] and [Shapiro et al., 2009] provide conditions for almost sure convergence in stochastic programming problems as we introduce in Section 2.6. The key requirements are that the objective function converges to its limiting function uniformly in $\mathbf{x}$ and that the optimal $\mathbf{x}^*$ is not on the boundary of the feasible space.

Other statistical properties have also been examined. [Aitchison and Silvey, 1958] provides conditions for asymptotic normality of maximum likelihood esti-

mators when the feasible region $\mathcal{X}$ of $\mathbf{x}$ contains equality constraints. The case with inequality constraints is less straightforward; [Shapiro \[1989\]](#) demonstrates that asymptotic normality of $\hat{\mathbf{x}}_n$ does not hold with inequality constraints if $\mathbf{x}^*$ is on the edge of the parameter space. [Cheng and Huang \[2010\]](#) provide conditions for bootstrap consistency for both $\hat{\theta}_n$ and $\hat{\mathbf{x}}_n$.

Nomenclature

Acronyms

CCGT	Combined cycle gas turbine
CLP	Continuous linear program
CO ₂ e	CO ₂ equivalent contribution (emissions)
IID	Independent and identically distributed
LDC	Load duration curve
MILP	Mixed integer linear program
PSM	Power system model

Mathematics

$:=$	Equal by definition
$ \mathcal{S} $	Cardinality of (number of elements in) set $\mathcal{S}$
$(\phi_t)_{t \in \mathcal{T}}$	Time series of vector ϕ_t across index set $\mathcal{T}$

Power system models

Indices & Sets

$i \in \mathcal{I}$	Generation technology
$r \in \mathcal{R}$	Region
$t \in \mathcal{T}$	Time step

Parameters

C_i^{gen}	Annualised generation install cost, technology i [£/MWyr]
$C_{r,r'}^{\text{tr}}$	Annualised transmission install cost, region r to r' [£/MWyr]
C^{sto}	Annualised storage energy install cost [£/MWhyr]
F_i	Generation cost, technology i [£/MWh]
E_i	Carbon emissions, technology i [$\text{kg CO}_2\text{e/MWh}$]
e^{sto}	Storage (dis)charge efficiency [$\in [0, 1]$]
l^{sto}	Storage self-loss [$1/\text{hr}$]

Time series

$d_{r,t}$	Demand, region r , time t [MWh]
$\lambda_{i,r,t}$	Generation potential, technology i , region r , time t [$\in [0, 1]$]
ξ_t	Time series values, time t

Decision variables

$\text{cap}_{i,r}^{\text{gen}}$	Generation capacity, technology i , region r [MW]
$\text{cap}_{r,r'}^{\text{tr}}$	Transmission capacity, region r to r' [MW]
$\text{cap}_r^{\text{sto}}$	Storage energy capacity, region r [MWh]
$\text{gen}_{i,r,t}$	Generation, technology i , region r , time t [MWh]
$\text{tr}_{r,r',t}$	Transmission, region r to r' , time t [MWh]
$\text{ch}_{r,t}$	Storage charging, region r , time t [MWh]
$\text{sto}_{r,t}$	Storage energy level, region r , time t [MWh]
$\mathbf{D}$	Power system design
$\mathbf{O}_t$	Power system operation, time t

Bibliography

- J. Aitchison and S. D. Silvey. Maximum Likelihood Estimation of Parameters Subject To Restraints. *The Annals of Mathematical Statistics*, 29(3):813–828, 1958. doi:[10.1214/aoms/1177706538](https://doi.org/10.1214/aoms/1177706538)
- R. Alzbutas and E. Norvaisa. Uncertainty and sensitivity analysis for economic optimisation of new energy source in Lithuania. *Progress in Nuclear Energy*, 61:17–25, 2012. doi:[10.1016/j.pnucene.2012.06.006](https://doi.org/10.1016/j.pnucene.2012.06.006)
- F. Amorim, S. G. Simoes, G. Siggini, and E. Assoumou. Introducing climate variability in energy systems modelling. *Energy*, 206, 2020. doi:[10.1016/j.energy.2020.118089](https://doi.org/10.1016/j.energy.2020.118089)
- B. Bahl, T. Söhler, M. Hennen, and A. Bardow. Typical periods for two-stage synthesis by time-series aggregation with bounded error in objective function. *Frontiers in Energy Research*, 5(January):1–13, 2018. doi:[10.3389/fenrg.2017.00035](https://doi.org/10.3389/fenrg.2017.00035)
- A. A. Bazmi and G. Zahedi. Sustainable energy systems: Role of optimization modeling techniques in power generation and supply - A review. *Renewable and Sustainable Energy Reviews*, 15(8):3480–3500, 2011. doi:[10.1016/j.rser.2011.05.003](https://doi.org/10.1016/j.rser.2011.05.003)
- P. J. Bickel and D. A. Freedman. Some Asymptotic theory for the bootstrap. *The Annals of Statistics*, 9(6):1196–1217, 1981. doi:[10.1214/aos/1176345637](https://doi.org/10.1214/aos/1176345637)
- P. J. Bickel, F. Götzte, and W. R. Van Zwet. Resampling fewer than n observations: Gains, losses, and remedies for losses. *Statistica Sinica*, 7(1):1–31, 1997. doi:[10.1007/978-1-4614-1314-1_17](https://doi.org/10.1007/978-1-4614-1314-1_17)
- M. M. bin Othman, A. Mohamed, and A. Hussain. Determination of transmission reliability margin using parametric bootstrap technique. *IEEE Transactions on Power Systems*, 23(4):1689–1700, 2008. doi:[10.1109/TPWRS.2008.2004734](https://doi.org/10.1109/TPWRS.2008.2004734)
- J. R. Birge and F. Louveaux. *Introduction to stochastic programming*. Springer, 2 edition, 2011. ISBN 9781461402367. doi:[10.1007/978-1-4614-0237-4](https://doi.org/10.1007/978-1-4614-0237-4)
- H. C. Bloomfield, D. J. Brayshaw, L. C. Shaffrey, P. J. Coker, and H. E. Thornton. Quantifying the increasing sensitivity of power systems to climate variability. *Environmental Research Letters*, 11(12):124025, 2016. doi:[10.1088/1748-9326/11/12/124025](https://doi.org/10.1088/1748-9326/11/12/124025)
- H. C. Bloomfield, D. J. Brayshaw, and A. J. Charlton-Perez. Characterizing the winter meteorological drivers of the European electricity system using targeted circulation types. *Meteorological Applications*, 2019. doi:[10.1002/met.1858](https://doi.org/10.1002/met.1858)
- H. C. Bloomfield, D. J. Brayshaw, and A. Charlton-Perez. MERRA2 derived time series of European country-aggregate electricity demand, wind power generation and solar power generation. *University of Reading, Dataset*, 2020. doi:[10.17864/1947.239](https://doi.org/10.17864/1947.239)
- H. C. Bloomfield, P. L. M. Gonzalez, J. K. Lundquist, L. P. Stoop, A. Kies, J. Browell, R. Dargaville, M. De Felice, K. Gruber, A. P. Hilbers, M. Panteli, H. E. Thornton, J. Wohland, M. Zeyringer, and D. J. Brayshaw. The importance of weather and climate to energy systems: A workshop on next generation challenges in energy-climate modelling.

- Bulletin of the American Meteorological Society*, pages 1–23, 2021. doi:[10.1175/bams-d-20-0256.1](https://doi.org/10.1175/bams-d-20-0256.1).
- N. Bodini, J. K. Lundquist, D. Zardi, and M. Handschy. Year-to-year correlation, record length, and overconfidence in wind resource assessment. *Wind Energy Science*, 1(2): 115–128, 2016. doi:[10.5194/wes-1-115-2016](https://doi.org/10.5194/wes-1-115-2016).
- V. Bosetti, G. Marangoni, E. Borgonovo, L. Diaz Anadon, R. Barron, H. C. McJeon, S. Politis, and P. Friley. Sensitivity to energy technology costs: A multi-model comparison analysis. *Energy Policy*, 80:244–263, 2015. doi:[10.1016/j.enpol.2014.12.012](https://doi.org/10.1016/j.enpol.2014.12.012).
- C. Bothwell and B. F. Hobbs. Sensitivity of resource adequacy to sample error in systems with dependent renewable generation. In *2018 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS 2018)*, pages 1–6, 2018. ISBN 978-1-5386-3596-4. doi:[10.1109/PMAPS.2018.8440524](https://doi.org/10.1109/PMAPS.2018.8440524).
- S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 7 edition, 2009. ISBN 9780511804441. doi:[10.1017/CBO9780511804441](https://doi.org/10.1017/CBO9780511804441).
- R. Bryce, I. Losada Carreño, A. Kumler, B. M. Hodge, B. Roberts, and C. Brancucci Martinez-Anido. Consequences of neglecting the interannual variability of the solar resource: A case study of photovoltaic power among the Hawaiian Islands. *Solar Energy*, 167(January):61–75, 2018. doi:[10.1016/j.solener.2018.03.085](https://doi.org/10.1016/j.solener.2018.03.085).
- K. K. Cao, K. Von Krbek, M. Wetzel, F. Cebulla, and S. Schreck. Classification and evaluation of concepts for improving the performance of applied energy system optimization models. *Energies*, 12(24), 2019. doi:[10.3390/en12244656](https://doi.org/10.3390/en12244656).
- X. Chen, Z. Y. Dong, K. Meng, Y. Xu, K. P. Wong, and H. W. Ngan. Electricity price forecasting with extreme learning machine and bootstrapping. *IEEE Transactions on Power Systems*, 27(4):2055–2062, 2012. doi:[10.1109/TPWRS.2012.2190627](https://doi.org/10.1109/TPWRS.2012.2190627).
- G. Cheng and J. Z. Huang. Bootstrap consistency for general semiparametric M-estimation. *Annals of Statistics*, 38(5):2884–2915, 2010. doi:[10.1214/10-AOS809](https://doi.org/10.1214/10-AOS809).
- Climate Action Tracker. Paris Agreement turning point. Technical report, Climate Action Tracker, 2020. URL <https://climateactiontracker.org/publications/global-update-paris-agreement-turning-point/>.
- S. Collins, J. P. Deane, K. Poncelet, E. Panos, R. C. Pietzcker, E. Delarue, and B. P. Ó Gallachóir. Integrating short term variations of the power system into integrated energy system models: A methodological review. *Renewable and Sustainable Energy Reviews*, 76:839–856, 2017. doi:[10.1016/j.rser.2017.03.090](https://doi.org/10.1016/j.rser.2017.03.090).
- S. Collins, P. Deane, B. Ó Gallachóir, S. Pfenninger, and I. Staffell. Impacts of inter-annual wind and solar variations on the European power system. *Joule*, 2(10):2076–2090, 2018. doi:[10.1016/j.joule.2018.06.020](https://doi.org/10.1016/j.joule.2018.06.020).
- Committee on Climate Change. The Sixth Carbon Budget: The UK’s path to Net Zero. Technical report, Committee on Climate Change, 2020. URL <https://www.theccc.org.uk/publication/sixth-carbon-budget/>.

- M. T. Craig, I. Losada Carreño, M. Rossol, B. M. Hodge, and C. Brancucci. Effects on power system operations of potential changes in wind and solar generation potential under climate change. *Environmental Research Letters*, 14(3):34014, 2019. doi:[10.1088/1748-9326/aaf93b](https://doi.org/10.1088/1748-9326/aaf93b).
- G. B. Dantzig and P. W. Glynn. Parallel processors for planning under uncertainty. *Annals of Operations Research*, 22(1):1–21, 1990. doi:[10.1007/BF02023045](https://doi.org/10.1007/BF02023045).
- A. Davison and D. V. Hinkley. *Bootstrap methods and their application*. Cambridge University Press, 1997. ISBN 978-0-521-57471-6. doi:[10.2307/1271471](https://doi.org/10.2307/1271471).
- P. De Guibert, B. Shirizadeh, and P. Quirion. Variable time-step: A method for improving computational tractability for energy system models with long-term storage. *Energy*, 213:119024, 2020. doi:[10.1016/j.energy.2020.119024](https://doi.org/10.1016/j.energy.2020.119024).
- C. De Jonghe, E. Delarue, R. Belmans, and W. D’haeseleer. Determining optimal electricity technology mix with high level of wind power penetration. *Applied Energy*, 88(6):2231–2238, 2011. doi:[10.1016/j.apenergy.2010.12.046](https://doi.org/10.1016/j.apenergy.2010.12.046).
- F. J. De Sisternes and M. D. Webster. Optimal selection of sample weeks for approximating the net load in generation planning problems. *MIT ESD Working Paper Series*, 0, 2013.
- J. DeCarolis, H. Daly, P. Dodds, I. Keppo, F. Li, W. McDowall, S. Pye, N. Strachan, E. Trutnevyte, W. Usher, M. Winning, S. Yeh, and M. Zeyringer. Formalizing best practice for energy system optimization modelling. *Applied Energy*, 194:184–198, 2017. doi:[10.1016/j.apenergy.2017.03.001](https://doi.org/10.1016/j.apenergy.2017.03.001).
- Department for Business Energy and Industrial Strategy. Electricity Generation Costs (November 2016). Technical report, Department for Business, Energy and Industrial Strategy, 2016. URL https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/566567/BEIS_Electricity_Generation_Cost_Report.pdf.
- B. Efron. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7(1):1–26, 1979. doi:[10.1214/aos/1176344552](https://doi.org/10.1214/aos/1176344552).
- Elexon. Value of lost load review process. Technical report, Elexon, 2018. URL https://www.elexon.co.uk/wp-content/uploads/2017/09/33_278_10_VoLL-Review-Process-Paper-v1.0.pdf.
- S. Fan and R. J. Hyndman. Short-term load forecasting based on a semi-parametric additive model. *IEEE Transactions on Power Systems*, 27(1):134–141, 2012. doi:[10.1109/TPWRS.2011.2162082](https://doi.org/10.1109/TPWRS.2011.2162082).
- D. Z. Fitiwi, F. De Cuadra, L. Olmos, and M. Rivier. A new approach of clustering operational states for power network expansion planning problems dealing with RES (renewable energy source) generation operational variability and uncertainty. *Energy*, 90:1360–1376, 2015. doi:[10.1016/j.energy.2015.06.078](https://doi.org/10.1016/j.energy.2015.06.078).
- J. Forrest, T. Ralphs, S. Vigerske, H. G. Santos, L. Hafer, B. Kristjansson, J. P. Fasano, E. Straver, M. Lubin, R. Lougee, H. I. Gassmann, and M. Saltzman. COIN-OR Branch-and-Cut Solver, version 2.10.5. mar 2020. doi:[10.5281/zenodo.3700700](https://doi.org/10.5281/zenodo.3700700).

- P. Gabrielli, M. Gazzani, E. Martelli, and M. Mazzotti. Optimal design of multi-energy systems with seasonal storage. *Applied Energy*, 219:408–424, 2018. doi:[10.1016/j.apenergy.2017.07.142](https://doi.org/10.1016/j.apenergy.2017.07.142).
- P. W. Glynn. Importance sampling for Monte Carlo estimation of quantiles. *Proceedings of the Second International Workshop on Mathematical Methods in Stochastic Simulation and Experimental Design*, pages 180–185, 1996.
- S. Goderbauer, M. Comis, and F. J. Willamowski. The synthesis problem of decentralized energy systems is strongly NP-hard. *Computers and Chemical Engineering*, 124:343–349, 2019. doi:[10.1016/j.compchemeng.2019.02.002](https://doi.org/10.1016/j.compchemeng.2019.02.002).
- S. Gonzato, K. Bruninx, and E. Delarue. Long term storage in generation expansion planning models with a reduced temporal scope. *Applied Energy*, 298:117168, 2021. doi:[10.1016/j.apenergy.2021.117168](https://doi.org/10.1016/j.apenergy.2021.117168).
- P. Härtel, M. Kristiansen, and M. Korpås. Assessing the impact of sampling and clustering techniques on offshore grid expansion planning. *Energy Procedia*, 137:152–161, 2017. doi:[10.1016/j.egypro.2017.10.342](https://doi.org/10.1016/j.egypro.2017.10.342).
- A. P. Hilbers, D. J. Brayshaw, and A. Gandy. Importance subsampling: improving power system planning under climate-based uncertainty. *Applied Energy*, 251:113114, 2019. doi:[10.1016/j.apenergy.2019.04.110](https://doi.org/10.1016/j.apenergy.2019.04.110).
- A. P. Hilbers, D. J. Brayshaw, and A. Gandy. Importance subsampling for power system planning under multi-year demand and weather uncertainty. In *2020 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS 2020)*, pages 1–6. IEEE, 2020. ISBN 978-1-7281-2822-1. doi:[10.1109/PMAPS47429.2020.9183591](https://doi.org/10.1109/PMAPS47429.2020.9183591).
- A. P. Hilbers, D. J. Brayshaw, and A. Gandy. Efficient quantification of the impact of demand and weather uncertainty in power system models. *IEEE Transactions on Power Systems*, 36(3):1771–1779, 2021. doi:[10.1109/TPWRS.2020.3031187](https://doi.org/10.1109/TPWRS.2020.3031187).
- A. P. Hilbers, D. J. Brayshaw, and A. Gandy. Reducing climate risk in energy system planning: a posteriori time series aggregation for models with storage. *arXiv preprint arXiv:2210.08351*, 2022.
- M. Hoffmann, L. Kotzur, D. Stolten, and M. Robinius. A review on time series aggregation methods for energy system models. *Energies*, 13(3):641, 2020. doi:[10.3390/en13030641](https://doi.org/10.3390/en13030641).
- T. Homem-de Mello and G. Bayraksan. Monte Carlo sampling-based methods for stochastic optimization. *Surveys in Operations Research and Management Science*, 19(1):56–85, 2014. doi:[10.1016/j.sorms.2014.05.001](https://doi.org/10.1016/j.sorms.2014.05.001).
- P. J. Huber. Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*, 35(1):73–101, 1964. doi:[10.1214/aoms/1177703732](https://doi.org/10.1214/aoms/1177703732).
- G. Infanger. Monte Carlo (importance) sampling within a benders decomposition algorithm for stochastic linear programs. *Annals of Operations Research*, 39(1):69–95, 1992. doi:[10.1007/BF02060936](https://doi.org/10.1007/BF02060936).

- J. D. Jenkins, E. N. Mayfield, E. D. Larson, S. W. Pacala, and C. Greig. Mission net-zero America: The nation-building path to a prosperous, net-zero emissions economy. *Joule*, 5(11):2755–2761, 2021. doi:[10.1016/j.joule.2021.10.016](https://doi.org/10.1016/j.joule.2021.10.016).
- S. Kamalinia and M. Shahidehpour. Generation expansion planning in wind-thermal power systems. *IET Generation, Transmission and Distribution*, 4(8):940–951, 2010. doi:[10.1049/iet-gtd.2009.0695](https://doi.org/10.1049/iet-gtd.2009.0695).
- S. Kamalinia, M. Shahidehpour, and A. Khodaei. Security-constrained expansion planning of fast-response units for wind integration. *Electric Power Systems Research*, 81(1):107–116, 2011. doi:[10.1016/j.epsr.2010.07.017](https://doi.org/10.1016/j.epsr.2010.07.017).
- S. Kennedy. Wind power planning: assessing long-term costs and benefits. *Energy Policy*, 33(13):1661–1675, 2005. doi:[10.1016/j.enpol.2004.02.004](https://doi.org/10.1016/j.enpol.2004.02.004).
- A. Khosravi, S. Nahavandi, and D. Creighton. Prediction intervals for short-term wind farm power generation forecasts. *IEEE Transactions on Sustainable Energy*, 4(3):602–610, 2013. doi:[10.1109/TSTE.2012.2232944](https://doi.org/10.1109/TSTE.2012.2232944).
- L. Kotzur, P. Markewitz, M. Robinius, and D. Stolten. Impact of different time series aggregation methods on optimal energy system design. *Renewable Energy*, 117:474–487, 2018a. doi:[10.1016/j.renene.2017.10.017](https://doi.org/10.1016/j.renene.2017.10.017).
- L. Kotzur, P. Markewitz, M. Robinius, and D. Stolten. Time series aggregation for energy system design: Modeling seasonal storage. *Applied Energy*, 213:123–135, 2018b. doi:[10.1016/j.apenergy.2018.01.023](https://doi.org/10.1016/j.apenergy.2018.01.023).
- L. Kotzur, L. Nolting, M. Hoffmann, T. Groß, A. Smolenko, J. Priesmann, H. Büsing, R. Beer, F. Kullmann, B. Singh, A. Praktiknjo, D. Stolten, and M. Robinius. A modeler’s guide to handle complexity in energy systems optimization. *Advances in Applied Energy*, 4:100063, 2021. doi:[10.1016/j.adapen.2021.100063](https://doi.org/10.1016/j.adapen.2021.100063).
- A. Kumler, I. L. Carreño, M. T. Craig, B. M. Hodge, W. Cole, and C. Brancucci. Inter-annual variability of wind and solar electricity generation and capacity values in Texas. *Environmental Research Letters*, 14(4):44032, 2019. doi:[10.1088/1748-9326/aaf935](https://doi.org/10.1088/1748-9326/aaf935).
- A. Lawson, M. Goldstein, and C. J. Dent. Bayesian framework for power network planning under uncertainty. *Sustainable Energy, Grids and Networks*, 7:47–57, 2016. doi:[10.1016/j.segan.2016.05.003](https://doi.org/10.1016/j.segan.2016.05.003).
- S. P. Lloyd. Least Squares Quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137, 1982. doi:[10.1109/TIT.1982.1056489](https://doi.org/10.1109/TIT.1982.1056489).
- P. Lopion, P. Markewitz, M. Robinius, and D. Stolten. A review of current challenges and trends in energy systems modeling. *Renewable and Sustainable Energy Reviews*, 96:156–166, 2018. doi:[10.1016/j.rser.2018.07.045](https://doi.org/10.1016/j.rser.2018.07.045).
- R. Loulou, G. Goldstein, and K. Noble. Documentation for the MARKAL Family of Models. *International Energy Agency*, (October):1–389, 2004. URL http://www.iea-etsap.org/web/Mrk1Doc-I_StdMARKAL.pdf.

- H. Lund, F. Arler, P. A. Østergaard, F. Hvelplund, D. Connolly, B. V. Mathiesen, and P. Karnøe. Simulation versus optimisation: Theoretical positions in energy system modelling. *Energies*, 10(7):1–17, 2017. doi:[10.3390/en10070840](https://doi.org/10.3390/en10070840).
- J. H. Merrick. On representation of temporal variability in electricity capacity planning models. *Energy Economics*, 59:261–274, 2016. doi:[10.1016/j.eneco.2016.08.001](https://doi.org/10.1016/j.eneco.2016.08.001).
- N. Metropolis and S. Ulam. The Monte Carlo Method. *Journal of the American Statistical Association*, 44(247):335–341, 1949.
- P. Nahmmacher, E. Schmid, L. Hirth, and B. Knopf. Carpe diem: a novel approach to select representative days for long-term power system modeling. *Energy*, 112:430–442, 2016. doi:[10.1016/j.energy.2016.06.081](https://doi.org/10.1016/j.energy.2016.06.081).
- National Grid. Future Energy Scenarios. Technical report, National Grid, 2021. URL <https://www.nationalgrideso.com/future-energy/future-energy-scenarios/fes-2021>.
- P. Parpas, B. Ustun, M. Webster, and Q. K. Tran. Importance sampling in stochastic programming: a Markov Chain Monte Carlo approach. *INFORMS Journal on Computing*, 27(2):358–377, 2015. doi:[10.1287/ijoc.2014.0630](https://doi.org/10.1287/ijoc.2014.0630).
- S. Pfenninger. Dealing with multiple decades of hourly wind and PV time series in energy models: a comparison of methods to reduce time resolution and the planning implications of inter-annual variability. *Applied Energy*, 197:1–13, 2017. doi:[10.1016/j.apenergy.2017.03.051](https://doi.org/10.1016/j.apenergy.2017.03.051).
- S. Pfenninger and B. Pickering. Calliope: a multi-scale energy systems modelling framework. *Journal of Open Source Software*, 3(29):825, 2018. doi:[10.21105/joss.00825](https://doi.org/10.21105/joss.00825).
- S. Pfenninger, A. Hawkes, and J. Keirstead. Energy systems modeling for twenty-first century energy challenges. *Renewable and Sustainable Energy Reviews*, 33:74–86, 2014. doi:[10.1016/j.rser.2014.02.003](https://doi.org/10.1016/j.rser.2014.02.003).
- S. Pineda and J. M. Morales. Chronological time-period clustering for optimal capacity expansion planning with storage. *IEEE Transactions on Power Systems*, 33(6):7162–7170, 2018. doi:[10.1109/TPWRS.2018.2842093](https://doi.org/10.1109/TPWRS.2018.2842093).
- K. Poncetlet, E. Delarue, D. Six, J. Duerinck, and W. D’haeseleer. Impact of the level of temporal and operational detail in energy-system planning models. *Applied Energy*, 162: 631–643, 2016. doi:[10.1016/j.apenergy.2015.10.100](https://doi.org/10.1016/j.apenergy.2015.10.100).
- K. Poncetlet, H. Hoschle, E. Delarue, A. Virag, and W. Drhaeseleer. Selecting representative days for capturing the implications of integrating intermittent renewables in generation expansion planning problems. *IEEE Transactions on Power Systems*, 32(3):1936–1948, 2017. doi:[10.1109/TPWRS.2016.2596803](https://doi.org/10.1109/TPWRS.2016.2596803).
- S. Pye, N. Sabio, and N. Strachan. An integrated systematic analysis of uncertainties in UK energy transition pathways. *Energy Policy*, 87:673–684, 2015. doi:[10.1016/j.enpol.2014.12.031](https://doi.org/10.1016/j.enpol.2014.12.031).

- P. Ravesteyn, G. van der Schrier, R. Haarsma, R. Scheele, and M. van den Broek. Vulnerability of European intermittent renewable energy supply to climate change and climate variability. *Renewable and Sustainable Energy Reviews*, 97:497–508, 2018. doi:[10.1016/j.rser.2018.08.057](https://doi.org/10.1016/j.rser.2018.08.057).
- M. M. Rienecker, M. J. Suarez, R. Gelaro, R. Todling, J. Bacmeister, E. Liu, M. G. Bosilovich, S. D. Schubert, L. Takacs, G. K. Kim, S. Bloom, J. Chen, D. Collins, A. Conaty, A. Da Silva, W. Gu, J. Joiner, R. D. Koster, R. Lucchesi, A. Molod, T. Owens, S. Pawson, P. Pegion, C. R. Redder, R. Reichle, F. R. Robertson, A. G. Ruddick, M. Sienkiewicz, and J. Woollen. MERRA: NASA’s modern-era retrospective analysis for research and applications. *Journal of Climate*, 24(14):3624–3648, 2011. doi:[10.1175/JCLI-D-11-00015.1](https://doi.org/10.1175/JCLI-D-11-00015.1).
- H. K. Ringkjøb, P. M. Haugan, and I. M. Solbrekke. A review of modelling tools for energy and electricity systems with large shares of variable renewables. *Renewable and Sustainable Energy Reviews*, 96:440–459, 2018. doi:[10.1016/j.rser.2018.08.002](https://doi.org/10.1016/j.rser.2018.08.002).
- S. Samsatli and N. J. Samsatli. A general spatio-temporal model of energy systems with a detailed account of transport and storage. *Computers and Chemical Engineering*, 80:155–176, 2015. doi:[10.1016/j.compchemeng.2015.05.019](https://doi.org/10.1016/j.compchemeng.2015.05.019).
- M. R. Shaner, S. J. Davis, N. S. Lewis, and K. Caldeira. Geophysical constraints on the reliability of solar and wind power in the United States. *Energy and Environmental Science*, 11(4):914–925, 2018. doi:[10.1039/c7ee03029k](https://doi.org/10.1039/c7ee03029k).
- A. Shapiro. Asymptotic properties of statistical estimators in stochastic programming. *The Annals of Statistics*, 17(2):841–858, 1989. doi:[10.1214/aos/1176347146](https://doi.org/10.1214/aos/1176347146).
- A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on stochastic programming: modeling and theory*. MPS/SIAM Series on Optimization, 2009. ISBN 9780898716870. doi:[10.1137/1.9780898718751](https://doi.org/10.1137/1.9780898718751).
- K. Singh. On the Asymptotic Accuracy of Efron’s Bootstrap. *The Annals of Statistics*, 9(6):1187–1195, 1981. doi:[10.1214/aos/1176345636](https://doi.org/10.1214/aos/1176345636).
- A. Soroudi and T. Amraee. Decision making under uncertainty in energy systems: state of the art. *Renewable and Sustainable Energy Reviews*, 28:376–384, 2013. doi:[10.1016/j.rser.2013.08.039](https://doi.org/10.1016/j.rser.2013.08.039).
- I. Staffell. Drax Electric Insights - Methodology and Sources. URL <https://reports.electricinsights.co.uk/methodology/>.
- I. Staffell and S. Pfenninger. The increasing impact of weather on electricity supply and demand. *Energy*, 145:65–78, 2018. doi:[10.1016/j.energy.2017.12.051](https://doi.org/10.1016/j.energy.2017.12.051).
- S. Stoft. *Power System Economics*. Wiley-IEEE Press, 2002. ISBN 978-0-471-15040-4. doi:[10.1109/9780470545584](https://doi.org/10.1109/9780470545584).
- N. Strachan and P.-H. Li. Energy Modelling in the UK Briefing Paper 1: The Modelling Landscape. Technical report, 2021. URL https://d2e1qxpsswcpgz.cloudfront.net/uploads/2021/08/UKERC_BN_Modelling-strengths-weaknesses.pdf.

- M. Sun, F. Teng, X. Zhang, G. Strbac, and D. Pudjianto. Data-Driven Representative Day Selection for Investment Decisions: A Cost-Oriented Approach. *IEEE Transactions on Power Systems*, 34(4):2925–2936, 2019. doi:[10.1109/TPWRS.2019.2892619](https://doi.org/10.1109/TPWRS.2019.2892619).
- H. Teichgraeber and A. R. Brandt. Time-series aggregation for the optimization of energy systems: Goals, challenges, approaches, and opportunities. *Renewable and Sustainable Energy Reviews*, 157:111984, 2022. doi:[10.1016/j.rser.2021.111984](https://doi.org/10.1016/j.rser.2021.111984).
- H. Teichgraeber, L. E. Küpper, and A. R. Brandt. Designing reliable future energy systems by iteratively including extreme periods in time-series aggregation. *Applied Energy*, 304: 117696, 2021. doi:[10.1016/j.apenergy.2021.117696](https://doi.org/10.1016/j.apenergy.2021.117696).
- D. A. Tejada-Arango, M. Domeshek, S. Wogrin, and E. Centeno. Enhanced representative days and system states modeling for energy storage investment analysis. *IEEE Transactions on Power Systems*, 33(6):6534–6544, 2018. doi:[10.1109/TPWRS.2018.2819578](https://doi.org/10.1109/TPWRS.2018.2819578).
- J. Timmerman, M. Hennen, A. Bardow, P. Lodewijks, L. Vandevelde, and G. Van Eetvelde. Towards low carbon business park energy systems: a holistic techno-economic optimisation model. *Energy*, 125:747–770, 2017. doi:[10.1016/j.energy.2017.02.081](https://doi.org/10.1016/j.energy.2017.02.081).
- S. T. Tokdar and R. E. Kass. Importance sampling: A review. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(1):54–60, 2010. doi:[10.1002/wics.56](https://doi.org/10.1002/wics.56).
- I. Tsiropoulos, W. Nijs, D. Tarvydas, and P. Ruiz. Towards net-zero emissions in the EU energy system by 2050. Technical report, European Commission Joint Research Centre, 2020. doi:[10.2760/081488](https://doi.org/10.2760/081488).
- W. W. Tso, C. D. Demirhan, C. F. Heuberger, J. B. Powell, and E. N. Pistikopoulos. A hierarchical clustering decomposition algorithm for optimizing renewable power systems with storage. *Applied Energy*, 270:115190, 2020. doi:[10.1016/j.apenergy.2020.115190](https://doi.org/10.1016/j.apenergy.2020.115190).
- B. van der Heijde, A. Vandermeulen, R. Salenbien, and L. Helsen. Representative days selection for district energy system optimisation: a solar district heating system with seasonal storage. *Applied Energy*, 248:79–94, 2019. doi:[10.1016/j.apenergy.2019.04.030](https://doi.org/10.1016/j.apenergy.2019.04.030).
- A. Van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 2014. ISBN 978-0-521-46903-2. doi:[10.1515/9783110250282](https://doi.org/10.1515/9783110250282).
- B. van Zuijlen, W. Zappa, W. Turkenburg, G. van der Schrier, and M. van den Broek. Cost-optimal reliable power generation in a deep decarbonisation future. *Applied Energy*, 253:113587, 2019. doi:[10.1016/j.apenergy.2019.113587](https://doi.org/10.1016/j.apenergy.2019.113587).
- A. Wald. Note on the consistency of maximum likelihood estimators. *The Annals of Mathematical Statistics*, 20(4):595–601, 1949. doi:[10.1214/aoms/1177729952](https://doi.org/10.1214/aoms/1177729952).
- S. W. Wallace and S. E. Fleten. Stochastic Programming Models in Energy. *Handbooks in Operations Research and Management Science*, 10(C):637–677, 2003. doi:[10.1016/S0927-0507\(03\)10010-2](https://doi.org/10.1016/S0927-0507(03)10010-2).
- C. Wan, Z. Xu, P. Pinson, Z. Y. Dong, and K. P. Wong. Probabilistic forecasting of wind power generation using extreme learning machine. *IEEE Transactions on Power Systems*, 29(3):1033–1044, 2014. doi:[10.1109/TPWRS.2013.2287871](https://doi.org/10.1109/TPWRS.2013.2287871).

- M. Welsch, M. Howells, M. Bazilian, J. F. DeCarolis, S. Hermann, and H. H. Rogner. Modelling elements of smart grids - enhancing the OSeMOSYS (Open Source Energy Modelling System) code. *Energy*, 46(1):337–350, 2012. doi:[10.1016/j.energy.2012.08.017](https://doi.org/10.1016/j.energy.2012.08.017)
- S. Wogrin. Time series aggregation for optimization: One-size-fits-all? *arXiv e-prints*, arXiv:2206(8):2–4, 2022.
- J. Wohland, M. Reyers, J. Weber, and D. Witthaut. More homogeneous wind conditions under strong climate change decrease the potential for inter-state balancing of electricity in Europe. *Earth System Dynamics*, 8(4):1047–1060, nov 2017. doi:[10.5194/esd-8-1047-2017](https://doi.org/10.5194/esd-8-1047-2017)
- J. Wohland, M. Reyers, C. Märker, and D. Witthaut. Natural wind variability triggered drop in German redispatch volume and costs from 2015 to 2016. *PLoS ONE*, 13(1):1, 2018. doi:[10.1371/journal.pone.0190707](https://doi.org/10.1371/journal.pone.0190707)
- M. Xu. *Analysis of market incentives on power system planning and operations in liberalised electricity markets*. PhD thesis, Durham University, 2016.
- X. Yue, S. Pye, J. DeCarolis, F. G. Li, F. Rogan, and B. Gallachóir. A review of approaches to uncertainty assessment in energy system optimization models. *Energy Strategy Reviews*, 21:204–217, 2018. doi:[10.1016/j.esr.2018.06.003](https://doi.org/10.1016/j.esr.2018.06.003)
- W. Zappa, M. Junginger, and M. van den Broek. Is a 100% renewable European power system feasible by 2050? *Applied Energy*, 233-234:1027–1050, 2019. doi:[10.1016/j.apenergy.2018.08.109](https://doi.org/10.1016/j.apenergy.2018.08.109)
- M. Zeyringer, J. Price, B. Fais, P. H. Li, and E. Sharp. Designing low-carbon power systems for Great Britain in 2050 that are robust to the spatiotemporal and inter-annual variability of weather. *Nature Energy*, 3(5):395–403, 2018. doi:[10.1038/s41560-018-0128-x](https://doi.org/10.1038/s41560-018-0128-x)
- Y. Zhang, V. Cheng, D. S. Mallapragada, J. Song, and G. He. A Model-Adaptive clustering-Based Time Aggregation Method for Low-Carbon Energy System Optimization. *IEEE Transactions on Sustainable Energy*, pages 1–11, 2022. doi:[10.1109/tste.2022.3199571](https://doi.org/10.1109/tste.2022.3199571)