When the Wind Blows

Vulnerability of Power Networks with High Renewable Penetration

by

Fergus Willsmore

A thesis submitted in partial fulfilment for the
degree of Master of Philosophy

THE UNIVERSITY
of ADELAIDE

Faculty of Engineering, Computer and Mathematical Science
School of Mathematical Sciences

August 2020

Declaration of Authorship

I, Fergus Willsmore, declare that this thesis titled, 'When the wind blows: vulnerability of power networks with high wind penetration' and the work presented in it are my own.

- I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.
- I give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.
- I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Signed: _____

Date: _____

The University of Adelaide

Abstract

Faculty of Engineering, Computer and Mathematical Science
School of Mathematical Sciences

Master of Philosophy

by Fergus Willsmore

Investigating the vulnerability of power networks gives us insight into the network elements that require reinforcement in order to maintain highly reliable power. The current state of climate has become of increasing concern, and has caused a rapid increase in renewable energy. Unlike conventional generation, renewable energy is highly dependent on the weather, potentially exposing network vulnerabilities to *cascading failure* and *network congestion*. A new model of Nesti *et al.* [38] investigates *emergent cascades* that arise from fluctuations in renewable energy. In particular, the authors use the *theory of large deviations* in order to rank the most-likely *initial line failures*. On the other hand, there are a number of studies that link frequent *congestions* to *high wind penetration*, and use *generation re-dispatch* in order to estimate *wind curtailment*. In this thesis, we first extend the initial emergent cascade model, by using the *power spectrum* to identify significant cycles in solar generation, and find that this decreases the probabilities of line failure. Then, we develop a new long-term congestion management model, which we use to investigate the impact of new wind farm connections on wind curtailment in the South Australian power network. We predict the power output of the new wind farms using a combination of linear regression, ARMA models and quantile regression. We find that the Mid-North has the largest amount of spare network capacity, and that transmission upgrades must coincide with the integration of new wind farms in the South East.

Acknowledgements

I would like to acknowledge my supervisors, Giang Nguyen, Gary Glonek and David Vowles for their guidance and feedback during the many meetings over the span of my degree. I particularly want to thank Giang for providing great assistance in the mammoth task of editing this thesis.

I also want to thank my parents, Jane and Ian, for their continual support at home, and my dog, Jet, for his company during isolation amid the COVID-19 pandemic.

Contents

List of Figures

List of Tables

Abbreviations

AC	Alternating Current
ACF	auto-correlation function
AEMO	Australian Electricity Market Operator
ARMA	Auto-Regressive Moving Average
BOM	Bureau of Meteorology
DC	Direct Current
GWh	gigawatt-hours
Hz	Hertz
ISP	Integrated System Plan
LODF	Line Outage Distribution Factor
MVA	megavolt ampere
MW	megawatt
NEM	National Electricity Market
OPF	Optimal Power Flow
PACF	partial auto-correlation function
POE	Probability of Exceedence
PTDF	Power Transmission Distribution Factor
pu	per-unit
PV	Photovoltaic
REZ	Renewable Energy Zone
SA	South Australia
SAPN	SA Power Networks
SARMA	Seasonal Auto-Regressive Moving Average
SCADA	Supervisory Control and Data Acquisition
SVC	Static VAR Compensator

Symbols

$\mathbb{N}$	natural numbers $0, 1, 2, \ldots$
$\mathbb{R}$	real numbers
$\mathbb{E}$	expectation operator
$\mathbb{P}$	probability measure
$\mathbf{0}$	column vector with all entries equal to 0
$\mathbf{1}$	column vector with all entries equal to 1
$\boldsymbol{e}_i$	column vector with the ith index equal to 1 and 0 elsewhere
$\mathbb{1}$	indicator function
J	a matrix with all entries equal to 1
ε	a small positive number
y^*	logit transform of the variable y
Re	real part of a complex number
Im	imaginary part of a complex number
$\exp(t)$	exponential function
$\gamma(t)$	auto-covariance function
$\rho(t)$	auto-correlation function
$M(t)$	moment generating function
$\upsilon(\omega)$	power spectrum of frequency ω
$\{\varepsilon_t\}_{t\in\mathbb{N}}$	white noise process
a_i	auto-regressive coefficient corresponding to lag $i \geq 1$
m_j	moving average coefficient corresponding to lag $j \geq 0$
$b_n(x)$	nth basis function of a natural cubic spline
$G(n, m)$	graph with n nodes and m edges
N	edge-vertex incidence matrix
B	diagonal susceptance matrix

L	weighted Laplacian
L^+	Moore-Penrose pseudo-inverse of the weighted Lapacian L
H	power transmission distribution factors
c_ℓ	capacity of line ℓ
C	diagonal inverse capacity matrix
R_{ij}	the effective resistance between bus i and bus j
$M_t^{(k)}$	the number of line failures in stage t of cascade k
g_k	total generation at bus k
g_k^{avail}	available generation at bus k
d_k	total load at bus k
$d_k^{\max}$	maximum load at bus k
$\boldsymbol{p}$	power injections
$\boldsymbol{p}^{(\ell)}$	the most-likely power injections that cause line ℓ to fail
$\boldsymbol{p}'$	power injections after re-dispatch
$\boldsymbol{\mu}$	nominal power injections
$\boldsymbol{\mu}_s$	nominal values of the stochastic power injections
$\Sigma_{\boldsymbol{p}}$	steady-state covariance matrix of the power injections
$\Sigma_{\boldsymbol{p}_s}$	steady-state covariance matrix of the stochastic power injections
$\boldsymbol{f}$	normalised power flows
$\boldsymbol{f}^{(\ell)}$	the most-likely power flows after line ℓ fails
$\boldsymbol{f}'$	power flows after re-dispatch
ν_ℓ	nominal power flow of line ℓ
σ_ℓ^2	variance of the power flow of line ℓ
δ_ℓ	decay rate of line ℓ
$\phi_k^{(\ell)}$	the change in power flow of branch k given branch ℓ fails
$N_n(\boldsymbol{\mu}, \Sigma)$	n-dimensional multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance Σ
$TN_n(\boldsymbol{\mu}, \Sigma)$	n-dimensional truncated multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance Σ
W_t^j	power output of the jth wind farm at time t
W_t^{j*}	power output of the jth wind farm at time t in the logit space
Φ_t^h	wind speed of weather station h at time t

Θ_t^h	categorical wind direction of weather station h at time t
η_t^*	fitted value of the average regression model at time t in the logit space
$\widehat{\eta}_t^*$	fitted value of the corrected average regression model at time t in the logit space
ζ_t^*	simulated residual at time t in the logit space
$Q_{\widehat{e}_t^{j*}\|\widehat{\eta}_t^*}(\tau)$	the τth conditional quantile function of $\widehat{e}_t^{j*}$ given $\widehat{\eta}_t^*$
$\psi_t^*(\zeta_t^*\|\widehat{\eta}_t^*)$	transformed simulated residual at time t in the logit space

Dedicated to my dog,
and best pal, Jet.

Chapter 1

Introduction

1.1 Problem description

In the modern-day, we use power directly or indirectly in every aspect of life. Power networks supply electricity to residential homes and major industries including agriculture and hospitality. Without a constant supply of electricity, food may spoil and critical medical equipment may fail. There are many other consequences including the impact on the economy, and hence maintaining a highly reliable power network is crucial.

Originally, power networks were developed to deliver electricity produced from conventional generators that burn fossil fuels. Unfortunately, when fossil fuels are burned, large amounts of carbon dioxide is released into the atmosphere which contributes to global warming. The current state of climate has become of increasing concern, and has caused a rapid increase in clean renewable energy. Unlike conventional generation, renewable energy is highly dependent on the weather, potentially exposing vulnerabilities in the network.

A reliable power network is able to supply the *demand* with a high degree of confidence and experience negligible loss of *load*, more commonly known as *blackouts*. Large-scale blackouts may leave a tremendous number of consumers, including residential homes and commercial businesses, without power for long periods of time. Typically, power is restored within a few hours but may potentially take days. Power losses of this magnitude have a significant financial impact on affected businesses and the network operators. The main mechanism that leads to large-scale blackouts is called *cascading failure*, which is a process of successive *line failures* caused by

dependencies in the *network topology*. Therefore, a model is needed in order to identify the network elements vulnerable to cascade and motivate solutions to reduce the risk of blackouts.

In 2011, an unfortunate mistake made by a technician caused the initial outage of a major transmission line near San Diego. Power redistribution resulted in a fast-moving large-scale blackout that affected southern California and nearby parts of Mexico. Approximately 2.7 million customers went without power for up to 12 hours. A year later, the cascade model of Bernstein *et al.* [4] identified the San Diego region to be the most vulnerable area to cascading failure. This highlights the importance of cascade modelling in order to identify network vulnerabilities.

Commonly, models analyse *classical cascades* which result from a random disturbance, such as equipment malfunction, technical mistake, or extreme weather. However, more recently, high renewable penetration has motivated models to treat wind and solar energy as stochastic resources. Nesti *et al.* [38] introduced a new model to investigate *emergent cascades* which arise from random fluctuations in renewable energy. The initial evidence demonstrates that emergent cascades have a higher *average line propagation* than classical cascades and therefore are more likely to produce fast-moving large-scale blackouts. In reality, cascades are rare events as heavily-loaded transmission lines are relieved through *congestion management*.

There are numerous aspects to maintaining reliable power, one of which is managing network *congestions* in order to operate in a *secure* state. A network element is *congested* when it is operating outside its physical stability limit, and secure operation, among other things, refers to a power network with no congestions. Congestion management consists of both *short-term* and *long-term* management. Short-term management concerns operating in a secure state now, and *long-term* management is concerned with having sufficient transmission capacity in order to secure operation over the coming years. Managing congestions in the short and long-term reduces the risk of blackouts but may significantly impact *wind penetration* - the percentage of electricity supplied by wind energy.

Many countries around the world have invested in wind energy, as it has become one of the cheapest sources of power. This has resulted in high wind penetration levels, as seen in Table 1.1.

	Country	Penetration
1.	Denmark	41%
2.	Ireland	28%
3.	Portugal	24%
4.	Germany	21%
5.	Spain	19%

Table 1.1: Top 5 countries with the highest wind penetration in 2018 [65].

High wind penetration levels exacerbate network congestion for two reasons: the available wind resource is not correlated with demand, and wind farm locations are clustered such that they share transmission capacity. Moreover, wind farms require a large amount of empty land in order to be efficient and therefore are located in rural areas. Typically, generator connections to rural regions of the network either have a lower transmission capacity, or rely on a limited number of high-capacity transmission lines to transport power to major loads. Geographical correlations in the wind resource produce frequent periods where wind power output is near installed capacity, and may congest the network.

In order to alleviate congestion, generation is *re-dispatched*. The re-dispatch solution may *curtail* the power output of at least one wind farm, referred to as *wind curtailment*. The amount of curtailed power is a quantity of interest, which inflates short-term congestion management costs and constrains wind penetration levels. Therefore, it is crucial to understand the dynamics that cause wind curtailment and identify the network elements that impose the constraint. A deeper understanding will lead to long-term solutions for appropriate integration of wind farms in order to ensure sufficient transmission capacity is available.

The South Australian (SA) power network is a particularly interesting case study due to the high wind penetration levels - almost 40% of SA's generation in 2019 was from wind energy [52]. Furthermore, there are several wind farm projects that have approval and are currently pending development. These new wind farm developments are likely to begin operating in the coming few years, dramatically increasing wind capacity in SA. However, for SA to be successful in supporting the resulting higher penetration levels, there needs to be sufficient understanding about future congestions that might arise.

1.2 The South Australian power network

The SA power network is part of a larger interconnected power network called the National Electricity Market (NEM). The NEM interconnects five states of Australia: Queensland, New South Wales (including the Australian Capital Territory), Victoria, Tasmania and South Australia. Each state has its own generation capacity and can import/export power via inter-connectors between states. The Australian Energy Market Operator (AEMO) manages the NEM, and is responsible for dispatching power, maintaining network stability, forecasting renewable resource, network maintenance and planning. Moreover, they have partnerships with the Bureau of Meteorology (BOM) and the Australian Renewable Energy Agency (ARENA).

The SA power network consists of a transmission network operated by ElectraNet, and a distribution network operated by SA Power Networks (SAPN). In SA, gas contributed 47.4%, wind 39.5%, rooftop PV 7.9%, diesel and small non-scheduled generation 0.8%, battery 0.2% and large-scale solar 0.03% of the 14,503 GWh of power generation in SA in the 2018-19 financial year [52]. Therefore, SA is a coal-free state as the last coal generator, Northern Power Station, was decommissioned in May 2016. In comparison, the NEM relies on coal generation (59.9%) and a fairly minor contribution from wind (6.2%) [3]. In the same period, SA imported 791 GWh of power from Victoria and exported 1,259 GWh, which is a significant improvement from the 2016-17 financial year where the net flows were 2,725 GWh of power imports [49]. It is clear that new wind farms since mid 2017 have stabilised the generation deficit that immediately followed the closure of the coal power station.

The increased amount of non-synchronous generation presents a new set of challenges in SA, one of which is managing low system strength. System strength relates to the ability of the power network to manage fluctuations in supply and demand while maintaining stable voltage levels [20]. SA's power network has a very high ratio of wind generation compared to the demand in the state and is capable of meeting, and historically has met, over 120% of its demand from wind generation alone. The system strength of the power network can be low during these periods, due to reduced levels of online conventional synchronous generation [49]. However, control systems such as battery storage and synchronous condensers help to manage this. Synchronous condensers operate in a similar way to a conventional generator with a rotating mass, and have been used since the beginning of the SA power network.

Another aspect of AEMO's system strength arrangements for SA involves the cur-

tailment of wind generation during periods of very high wind output. During the 2018-19 financial year, approximately 3.3% of South Australian wind generation was curtailed to maintain the power system within secure limits. This was an increase from 2.4% of wind generation curtailed in the previous year [52]. Key drivers for increased curtailment included record high wind generation and insufficient synchronous generation due to periods of relatively low prices [48].

As part of the network planning process, AEMO published in 2018 the Integrated System Plan (ISP) [47], which is a cost-based optimisation that forecasts the transmission network requirements of the NEM over the next 20 years. AEMO's modelling estimates that the additional transmission investment proposed in the 2018 ISP delivers savings of around $1.2 billion. The transmission investment that concerns SA includes improved system strength and a new inter-connector. The lack of spinning turbines from conventional generation in SA, means there is an immediate requirement for synchronous condensers to supply system strength and inertia to the region. The new inter-connector RiverLink proposes a 750 MW transfer capacity between South Australia and New South Wales and is currently under assessment by ElectraNet. Expected benefits include increased system strength in SA and enable the connection of renewable energy in the Riverland.

Furthermore, AEMO assessed 34 candidate renewable energy zones (REZ) across the NEM, including 9 in SA: South East, Riverland, Mid-North, Yorke Peninsula, Northern SA, Leigh Creek, Roxby Downs, Eastern and Western Eyre Peninsula. The assessment identifies the Mid-North for least-cost integration of additional wind capacity into the transmission network. A significant factor of least-cost integration is the REZ's spare network capacity defined by the MW value of additional generation that can be transported from the REZ to the required load centre [47]. The ISP reports considerable spare network capacity of 1000 MW in both the Mid-North and Northern SA, and 500 MW in the South East. New renewable developments will utilise spare network capacity for cost-effective wind integration.

Currently, there are a number of approved wind farm projects spread out across SA. Tilt Renewables have operated wind farms in SA since 2008 and are now in the development phase of another 300 MW wind farm which is situated east of Adelaide [56]. DP Energy plan to commence development of their renewable energy park in Northern SA in mid 2020, which will include a 210 MW wind farm [21]. More recently, RES Australia was granted development consent for a 185 MW wind farm located in the Mid-North [60]. Finally, REpower Australia have approval for a 600 MW wind farm located on the Yorke Peninsula that will deliver electricity direct

to Adelaide via an undersea cable. REpower Australia have already delivered 6 wind farms in SA, equating to over 500 MW of installed capacity [2]. These promising renewable projects will establish sufficient wind capacity in SA to achieve higher wind penetration levels.

1.3 Structure of the thesis

Chapter 2 consists of four main sections. Section 2.1 introduces the physical phenomena of electrical circuits and power network security and stability before deriving DC power flows and power redistribution. Section 2.2 provides a broad overview of cascade models. We first present *physical* models that analyse *classical cascades* caused by random disturbances and the *emergent cascade* model of Nesti *et al.* [38] which analyses cascades that emerge due to fluctuations in renewable energy. We then consider *abstract* models of Dobson *et al.* [16], which along with other work of Dobson *et al.* [18] motivate cascade model validation methods of Qi *et al.* [55]. Similarly, Section 2.3 provides a literature review of network congestion and wind curtailment with an emphasis on *short-term* and *long-term* congestion management. We first describe case studies of Joos and Staffel [29] and of Luo *et al.* [33] that analyse observed wind curtailment from *short-term* congestion management. We then introduce a re-dispatch model of Gu *et al.* [24] to quantify wind curtailment relief provided by *long-term* congestion management, and comment on more generalised re-dispatch models [12, 35, 63]. Finally, we conclude this chapter with statistical concepts related to time series and regression analysis which are necessary to understand our work.

The two main contributions of this thesis are: an improved emergent cascade model and a new long-term congestion management model. In Chapter 3, we first discuss in-depth, the *emergent cascade* model of Nesti *et al.* [38]. The authors develop time series models for wind and solar generation in order to estimate the *covariance* of power injections for the German power network. In Section 3.2, we produce diagnostic plots of the wind and solar models in order to determine goodness-of-fit. The diagnostics of the solar model motivate Section 3.3, where we follow the work of Boland [9] and extend the model to incorporate a deterministic *Fourier series*. We then demonstrate that the extension of the solar model improves the goodness-of-fit and increases the accuracy of the estimates for the failure probabilities. We conclude this section with a discussion of our main results and future work.

The long-term congestion management model quantifies the impact of additional

wind capacity on wind curtailment in SA and consists of three new components: a network model of SA, a wind power simulation and a short-term generation re-dispatch model. To that end, Chapter 4 presents a simplified open-access AC model of the SA power network. First, Section 4.1 outlines the *physical* model which consists of the network elements whose parameters are publicly available, and the network elements whose parameters are estimated or assumed. Then, in Section 4.2, we balance the time series of generation and demand to determine the *dynamic inputs* to the *physical* model. In Section 4.3, we present diagnostic plots of our AC model to ensure its capability of emulating network operation. Then, in Section 4.4, we perform a DC approximation to the AC model to obtain a closed-form analytical solution, and assess the validity of the assumptions using conditions presented by Purchala *et al.* [54]. We conclude with a visualisation of the power transmission distribution factors of the estimated DC network model.

In Chapter 5, we first investigate potential methods to simulate a time series of wind generation in SA. Then, in Section 5.2, we briefly introduce the dataset of wind speed and wind direction obtained from the BOM, and apply pre-processing steps. In Section 5.3 we develop a *mean prediction* model which estimates the expected level of wind generation from a series of *linear regressions*. Then, in Section 5.4, we develop an *error simulation* model which simulates a stochastic deviation from the *mean prediction*, in order to obtain a realistic time series of wind generation. We conclude by validating our wind power simulations against historical data, using performance metrics such as correlation and probability of exceedance.

In our final chapter, we first develop two optimisation models to perform short-term generation re-dispatch in order to estimate the amount of wind curtailment. Then, in Section 6.2, we bring together the network model from Chapter 4, the wind power simulation from Chapter 5 and the short-term re-dispatch to form the long-term congestion management model. Results from the model are presented in Section 6.3, where we compare our findings to the 2018 ISP and demonstrate our models potential to aid long-term planning decisions. We conclude this section with a discussion of the limitations of our methodology and areas of future research.

Finally, in Chapter 7 we conclude the thesis, commenting on our contributions, future work, and lessons learned.

Chapter 2

Background

In order to investigate cascading failure and network congestion, we represent a power network as a connected graph $G(n, m)$, where the n nodes represent *buses* and the m edges represent network *branches*. The buses represent substations, which among other things, connect transmission elements such as lines, power generators, and loads. A network branch is either a *transmission line* or a *transformer*. Transmission lines are designed to transport generated power across the network with minimal losses. High-voltage transmission lines are used over longer distances in the transmission network and transformers are designed to convert power between different voltage levels. A bus may have a *generator* or *load* connected. The generators convert a fuel resource (potentially renewable) into power, where it is then transported across the transmission network until it is extracted by a load. The loads extract power equal to their demand and can represent a distribution network or an industrial customer. An example power network with 5 buses and 5 branches is represented as a graph $G(5, 5)$ in Figure 2.1.

Furthermore, we require knowledge of the power flows. There are two distinct types of power dynamics, namely *direct current* (DC) and *alternating current* (AC). All power networks are AC but the DC model is sometimes used to approximate the power flows. We now present electrical circuit and power network background to derive the DC power flow model.

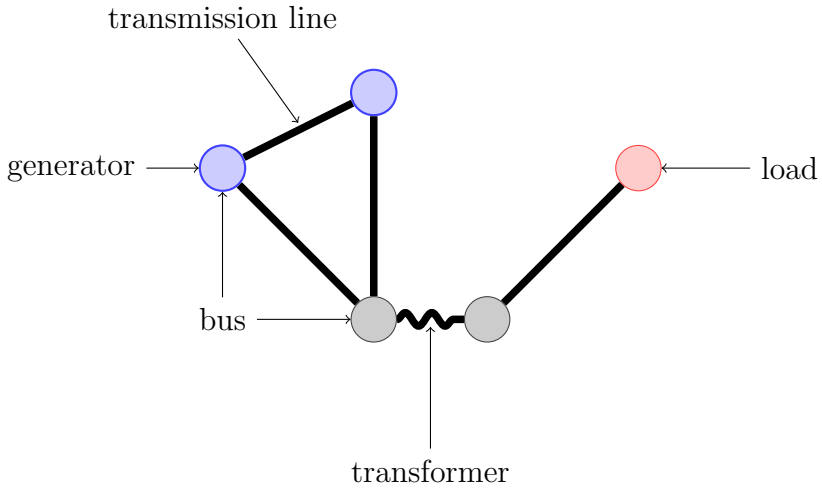


Figure 2.1: A basic example of a graph $G(5,5)$, which represents a power network transmission system. The generators inject power into the network at the blue buses, and the load extracts power from the network at the red bus. The grey buses have no generator or load attached and therefore the corresponding power injections are always zero.

2.1 Electric circuits and power networks

In this section, we present necessary material on power networks to understand this thesis, and for more details see [28]. First, consider the simpler DC circuit, *i.e.,* battery operated, where a fixed voltage is applied. The *voltage* (V) creates a potential electrical difference such that the electrons flow within a circuit with fixed direction toward the positive charge. This movement of electrons through a conductor is referred to as the *current* (I). *Electric power* (P) combines *current* and *voltage* such that

$$P = VI\,.$$

Ohm's law states that for any DC circuit the electric current is directly proportional to the voltage. This gives rise to the *resistance* (R) of the conductor as the proportionality constant such that

$$I = \frac{V}{R}\,.$$

An electrical network (or power network) consists of a series of closed circuits that obey *Kirchhoff's laws*. In particular, *Kirchoff's first law* states that for any node in an electrical circuit, the sum of currents flowing into a node is equal to the sum of currents flowing out of that node, and is known as the *conservation of charge*. In a power network the buses represent the nodes, as shown in Figure 2.1.

A power network has a specific topology which consists of the connection points between all the network elements such as transmission branches, generators and loads. Additionally, the network elements have parameter specifications that effect the power dynamics. For instance, generators have a maximal generation capacity that cannot be exceeded; branches have a maximal power carrying capacity that should not be exceeded; each bus is defined on a nominal voltage level; and there are resistive forces that create power losses in the network. Assuming all this information about a network is known, then given the power output of each generator and the demand of each load in the network, collectively known as the *power injections*, the DC power flows are solved using a system of linear equations. We derive the solution to the DC power flows in Section 2.1.3.

In an AC circuit, the voltage oscillates as a sinusoidal wave which causes the current to alternate directions. Real power networks utilise three-phase alternating current transmission for constant power transfer and an improved capacity-to-cost ratio. Three-phased power is achieved by running three wires in parallel each offset by 120 degrees. Specifically, an AC circuit has time-dependent current, $i(t)$, and voltage, $v(t)$, of the general form

$$v(t) = V_0 \cos(\omega t + \tau^v), \qquad i(t) = I_0 \cos(\omega t + \tau^i)\,, \text{ for } t \geq 0\,.$$

Here, V_0 and I_0 are the amplitudes of voltage and current, respectively, and τ^v and τ^i are the phase shifts of voltage and current, respectively. The parameter ω is the frequency of both voltage and current. Under stable conditions it is a system-wide parameter equal to $2\pi\omega_0$, where ω_0 is the nominal system frequency (50 Hz in Australia and 60 Hz in America). The relationship between sine waves and complex numbers is exploited to simplify the power flow analysis. Thus, defining the complex voltage and current phasor respectively as

$$V = \frac{V_0}{\sqrt{2}} e^{j\tau^v}, \qquad I = \frac{I_0}{\sqrt{2}} e^{j\tau^i}\,,$$

where $j = \sqrt{-1}$, we obtain the *active* (or real) *power flows*

$$p = \mathrm{Re}(V\overline{I})\,,$$

where Re indicates the real part, and $\overline{I}$ is the complex conjugate of the current, I. Consequently, the related quantities

$$q = \mathrm{Im}(V\overline{I}) \quad \text{and} \quad S = p + jq\,,$$

denote the *reactive power* and *apparent power*, respectively, where Im indicates the imaginary part. Therefore, the AC power flows are defined in terms of complex

numbers that involve active power and reactive power. Active power corresponds to power that is consumed by the loads. By the laws of electromagnetic induction, in order to supply active power, energy is consumed to create and maintain a magnetic field, called reactive power. Although reactive power does not contribute to the load, it is essential to maintain voltage levels across the power network.

Additionally, the complex behaviour gives rise to the *reactance*, X, which is the opposing force applied to a change in current or voltage, due to inductance or capacitance (energy stores in the form of magnetic and electric fields). The *impedance*, Z, is given by

$$Z = R + jX\,,$$

and the *admittance*, Y, is related such that

$$Y = Z^{-1} = G + jB\,,$$

where G and B are called the *conductance* and the *susceptance*, respectively. Given an electrical network with known AC power injections and impedance properties, the branch power flows are solutions to a system of quadratic equations. These are inherently more complex than the DC-equivalent system of linear equations. Fortunately, under normal stability conditions, the complex power flows can be approximated by linear equations to produce the DC approximation [5, 13, 38, 58]. Prior to defining the DC approximation, we highlight some important concepts applied in power systems engineering.

2.1.1 Per-unit system

In power systems analysis, the *per-unit system* (pu) is commonly used to express system quantities as a fraction of a defined base unit such that

$$P_{pu} = \frac{P}{P_{base}}\,, \quad V_{pu} = \frac{V}{V_{base}}\,, \quad Z_{pu} = \frac{Z}{Z_{base}} \quad \text{and} \quad Y_{pu} = \frac{Y}{Y_{base}}\,.$$

In system-wide calculations a common system base, P_{base}, is chosen. Typically, the base power is chosen to be $P_{base} = 100$ MVA to make the power quantities more convenient. Base voltage is commonly chosen to be the nominal voltage level of the bus. Power networks are typically interconnected systems that connect buses defined on different voltage levels. A transformer will connect adjacent buses and convert the power between the voltage level of the *sending* and *receiving* bus. The significance of the per-unit system is that the voltage on either side of the transformer is equal

to 1 pu. The base impedance and base admittance are calculated from electrical circuit laws as follows

$$\begin{aligned} Z_{base} &= \frac{V_{base}}{I_{base}} = \frac{V_{base}^2}{I_{base}V_{base}} = \frac{V_{base}^2}{P_{base}}, \\ Y_{base} &= \frac{1}{Z_{base}}. \end{aligned}$$

It is also simple to convert between a per-unit value defined on different power bases. One obvious need would be to convert a per-unit power flow solution back to the actual power flow solution. The new per-unit power flow solution P_{pu}^{new} with base P_{base}^{new} is given by

$$P_{pu}^{\text{new}} = P_{pu} \times \frac{P_{base}^{\text{new}}}{P_{base}}.$$

The per-unit power flows are converted back to the power flows using the identity base power $P_{base}^{\text{new}} = 1$ MVA. In Section 2.1.3 the calculations of the DC power flows use per-unit values, but we omit pu for notation purposes.

2.1.2 Security and stability

The notion of power system *security* and *stability* are closely related. Power system security refers to the state of grid operation, whilst stability refers to the power system's ability to regain a state of operating equilibrium after being subjected to a physical disturbance (*e.g.* line failure, transformer failure, generator failure, etc.). The power system can be operated in a secure, unsatisfactory or unstable state.

Operating in a secure state implies that the network parameters are within limits and will remain within limits if subjected to a single disturbance. Operating in an unsatisfactory state implies that the system is stable with some network parameters operating outside their limits and will most likely remain in an unsatisfactory state if subjected to a single disturbance. However, operating in an unsatisfactory state for long periods due to multiple disturbances can cause the system to move to an unstable state, where it is vulnerable to collapse.

Security is managed by the market operators who conduct the centrally-coordinated security-constrained optimised dispatch for the projected demand profile. The optimal dispatch schedules available generation to minimise operational costs subject to the network constraints. The network constraints correspond to transmission branch power flow capacities and bus voltage levels, which are continuously monitored in

real-time using system observation controllers such as Supervisory Control and Data Acquisition (SCADA).

Transmission elements are subject to faults that occur over time or caused by an exogenous disturbance. Such events can either immediately cause the network element to fail, *unplanned outage*, or are intentionally removed from service to perform make-safe maintenance, *planned outage*. As a result, operators incorporate a level of network redundancy in the dispatched generation solution in order to ensure that no load shedding occurs due to any failures, commonly known as *the $N-k$ problem* [6]. Typically, operators perform an $N-1$ level contingency analysis to ensure that failure of any element does not result in load shedding. However, this becomes convoluted for large interconnected systems with potentially concurrent elements offline for maintenance. In reality, operators need to consider the $N-k-p$ problem where there are k unplanned outages and p planned outages. Maintaining higher reliability is achieved with a substantial increase in operating cost.

Line failures cause system dynamics to deviate from a state of equilibrium. Large deviations will cause violations in operating limits and result in an unstable system vulnerable to collapse. We briefly summarise three key areas of power system stability: rotor-angle stability, frequency stability and voltage stability. *Rotor-angle stability* refers to the ability of the power network to remain in synchronism after being subjected to a disturbance. This requires sufficient synchronising torques between the rotors of synchronous generators, such that at equilibrium the *electromagnetic torque*, the force applied to the rotor from electromagnetic induction, is equal and opposite to the *mechanical torque*, the force applied to the rotor from burning fuel. When mechanical torque leads electromagnetic torque, the rotor is accelerating; conversely, when electromagnetic torque leads mechaninal torque, the rotor is decelerating.

Frequency stability concerns the ability of the power network to maintain balance between generation and demand following a disturbance. When demand exceeds the total generation, the system frequency falls below the nominal synchronous frequency, and increases above the nominal frequency when total generation exceeds demand. Unless balance is restored, the system frequency will continue to deviate from the nominal to an extent where synchronism will be lost. Frequency stability is managed by maintaining a low Rate-of-Change-of-Frequency (RoCoF) immediately following the disconnection of a large load or generator.

As a result of synchronous generation being replaced by asynchronous renewable generation, power networks have experienced a reduction in synchronous inertia [26].

Synchronous generators operate with high inertia as they consist of a rapidly moving rotor, which has momentum. On the other hand, asynchronous generators such as wind farms rotate at a variable speed that is much slower than the system frequency, and solar farms have no rotation. A disturbance to the power network applies a force of change to the system frequency that is countered by the synchronous inertia; therefore, it is important to maintain inertia levels for frequency stability.

Voltage stability refers to the ability of the power system to maintain balance between the supply and consumption of reactive power at all buses, in order to maintain bus voltages within prescribed limits.

2.1.3 DC power flows

The DC approximation to the AC power flows is reliable under the following conditions:

- **Voltage stability:** the *per-unit system* implies that $|V_k| \approx 1$ pu for each bus k, where V_k denotes the voltage at bus k.
- **Rotor-angle stability:** define the phase angle, θ_k, by the voltage phase shift τ^v at bus k. Under stable conditions, for each branch (k, h), the phase angle difference $\theta_k - \theta_h$ is sufficiently small so that $\sin(\theta_k - \theta_h) \approx \theta_k - \theta_h$.
- **Resistance is negligible:** for any branch (k, h), the resistance r_{kh} is much smaller than the reactance x_{kh}.

When the above conditions are satisfied, the approximate power flow along branch (k, h) is given by

$$f_{kh} = \frac{\theta_k - \theta_h}{x_{kh}} \,.$$

It is convenient to represent the power flow equations in matrix form, adopting from graph theory the edge-vertex incidence matrix and the weighted Laplacian matrix for a graph $G(n, m)$. The *edge-vertex incidence matrix* is defined by

$$Q_{(k,h),j} = \begin{cases} 1 & \text{for } j = k \,, \\ -1 & \text{for } j = h \,, \\ 0 & \text{otherwise} \,, \end{cases}$$

such that $Q \in \mathbb{R}^{m \times n}$ is an $m \times n$ matrix with row sums equal to zero. The $m \times m$ *susceptance* matrix is defined by

$$B = \text{diag}(\beta_1, \ldots, \beta_m) \,,$$

where $\beta_\ell = \beta_{kh} = -x_{kh}^{-1}$ represents the per-unit susceptance of branch $\ell = (k, h)$, commonly viewed as *edge weights*. The *weighted Laplacian matrix* is of the form $L = Q^\top BQ$ with elements

$$L_{kh} = \begin{cases} \sum\limits_{h:(k,h)\in G} \beta_{kh} & \text{for } k = h\,, \\ -\beta_{kh} & \text{for } k \neq h\,. \end{cases}$$

Under the DC model, the per-unit power flows are of the form

$$\widehat{\boldsymbol{f}} = BQ\boldsymbol{\theta}\,, \tag{2.1}$$

and the per-unit power injections are given by

$$\boldsymbol{p} = L\boldsymbol{\theta}\,,$$

where $\boldsymbol{\theta}$ is the column vector of phase angles. As the incidence Q has zero row sums, the Laplacian L is not invertible. Thus, another matrix, $\overline{L}$, obtained by using a *reference bus*, is commonly used. A reference, slack or swing bus has a fixed phase angle, $\theta_{\text{slack}} = 0$, and the power injection is given by the mismatch between generation and load. Without loss of generality, the phase angle of bus n is set to zero and is called the reference bus. The Laplacian L has rank $n-1$ and thus $\overline{L}$ is invertible such that

$$\boldsymbol{\theta} = \overline{L}^{-1}\boldsymbol{p}\,.$$

A real power network does not operate with a reference bus. Instead, network operators will adjust for the transmission losses using as many generator buses as possible in order to minimise system changes. Another method of Nesti *et al.* [38] uses the *Moore-Penrose pseudo-inverse* of the Laplacian L defined by

$$L^+ = \left(L + \frac{1}{n}J\right)^{-1} - \frac{1}{n}J\,, \tag{2.2}$$

where $J \in \mathbb{R}^{n\times n}$ is the $n \times n$ matrix with all entries equal to one. This method implicitly sets the average value of the phase angles to zero and will be used to define the notion of *effective resistance* in Section 2.1.4. When n is large, numerical errors may arise in (2.2) due to the matrix $L + \frac{1}{n}J$ being ill-conditioned. However, the matrix L^+ can be factored using a singular value decomposition. It directly follows that

$$\boldsymbol{\theta} = L^+\boldsymbol{p}\,. \tag{2.3}$$

Then, the substitution of Equation (2.3) into Equation (2.1) implies that the per-unit power flows on the system power base are given by

$$\widehat{\boldsymbol{f}} = BQL^+\boldsymbol{p}\,.$$

It is convenient to normalise the power flows by their corresponding branch capacities. This is a separate technique from the per-unit system which applies a system-wide scaling factor. Denote by $C \in \mathbb{R}^{m\times m}$ the diagonal matrix with diagonal elements c_ℓ^{-1} for every $\ell = 1, \ldots, m$, where c_ℓ is the capacity of branch ℓ, measured on the system power base, P_{base}. A linear transformation, defined by C, of the power flows results in the *normalised power flows*

$$\boldsymbol{f} = H\boldsymbol{p}\,, \tag{2.4}$$

where $H = CBQL^+ \in \mathbb{R}^{m\times n}$ is the matrix of power transmission distribution factors (PTDFs). Nesti *et al.* [38] assume that the failure of branch ℓ occurs when $|f_\ell| \geq 1$. After a branch failure, the power is redistributed throughout the altered network.

2.1.4 Power redistribution

In the literature it is common to use line outage distribution factors (LODFs) to calculate the redistribution of power. The following derivation of the LODFs is based on Cetinay *et al.* [13]. When a branch fails and is disconnected from operation, the power that was flowing across the disconnected branch redistributes throughout the network. The redistribution is governed by the LODFs and might cause other branches to become overloaded.

Given the failure event of branch ℓ, denote by $G^{(\ell)} = G(n, m\backslash\{\ell\})$ the graph that represents the network without branch ℓ. The power that flowed across branch ℓ redistributes throughout the altered network $G^{(\ell)}$. We are interested in finding the LODF coefficient $\phi_k^{(\ell)}$ that represents the change in power flow across branch k given the failure of branch ℓ. Denote by $f_k^{(\ell)}$ the power flow across branch k given the failure of branch ℓ. The new power flow on branch k after branch ℓ is disconnected is given by

$$f_k^{(\ell)} = f_k + \phi_k^{(\ell)} f_\ell^{(\ell)}\,,$$

for every $k \neq \ell$ and where $f_\ell^{(\ell)} = \pm 1$ depending on the power flow direction of branch ℓ. This is an efficient calculation that uses the previous power flow solution. The LODFs, $\phi_k^{(\ell)}$, are calculated using the notion of *effective resistance* (also known as *resistance distance* in graph theory), which is the resistance between two buses in the power network. Cetinay *et al.* [13] derive the effective resistance R_{ij} between bus i and bus j as follows:

$$R_{ij} = (L^+)_{ii} + (L^+)_{jj} - 2(L^+)_{ij}\,.$$

Consider the failure of branch $\ell = (i, j)$. The change in power flows can be expressed using the original power flow solution such that

$$\partial^{(\ell)} \boldsymbol{f} = CBQ\partial^{(\ell)}\boldsymbol{\theta}\,, \tag{2.5}$$

where $\partial^{(\ell)}$ denotes the change caused by the failure of branch ℓ. The new Laplacian $L^{(\ell)}$ of the altered network $G^{(\ell)}$ is given by

$$L^{(\ell)} = L - \beta_\ell Q_\ell Q_\ell^\top\,,$$

where β_ℓ represents the ℓth diagonal entry of the susceptance matrix, B, and Q_ℓ represents the ℓth column of the edge-vertex incidence matrix, Q. Intuitively, this says that the new Laplacian after branch ℓ fails is equal to the difference between the Laplacian before failure and the effect of branch ℓ. From *Meyer's relation* [36], the Moore-Penrose pseudo-inverse of the Laplacian $L^{(\ell)}$ is given by

$$L^{(\ell)+} = L^+ + \frac{L^+ Q_\ell \beta_\ell Q_\ell^\top L^+}{1 - \beta_\ell Q_\ell^\top L^+ Q_\ell}\,. \tag{2.6}$$

The change in the phase angles caused by the failure of branch ℓ is given by

$$\begin{aligned} \partial^{(\ell)}\boldsymbol{\theta} &= \partial^{(\ell)} L\boldsymbol{p} \\ &= \left(L^{(\ell)+} - L^+\right)\boldsymbol{p}\,. \end{aligned} \tag{2.7}$$

Power flow redistribution is immediate, hence the power injections remain the same after failure until re-dispatch. Substitute Equation (2.6) into (2.7) and then Equation (2.7) into (2.5). Then, the change in power flow along branch $k = (a, b)$ given the failure of branch $\ell = (i, j)$ is given by

$$\partial\, \boldsymbol{f}_k^{(\ell)} = \beta_k \frac{c_\ell}{c_k}\left(\frac{Q_k^\top L^+ Q_\ell}{1 - \beta_\ell Q_\ell^\top L^+ Q_\ell}\right) f_\ell^{(\ell)}\,,$$

where $f_\ell^{(\ell)} = c_\ell^{-1}\beta_\ell Q_\ell^\top L^+ P$. This can be simplified using the notion of effective resistance by noticing that

$$\begin{aligned} Q_k^\top L^+ Q_\ell &= (\boldsymbol{e}_a - \boldsymbol{e}_b)^\top L^+ (\boldsymbol{e}_i - \boldsymbol{e}_j) \\ &= (L^+)_{ai} - (L^+)_{aj} + (L^+)_{bi} - (L^+)_{bi} \\ &= \frac{1}{2}(R_{aj} - R_{ai} + R_{bi} - R_{bj})\,, \end{aligned}$$

where $\boldsymbol{e}_i$ is a unit vector equal to one in the ith index and zero elsewhere. Then, the LODF coefficient $\phi_k^{(\ell)}$ that represents the power change along branch k given branch ℓ fails is given by

$$\phi_k^{(\ell)} = \beta_k \frac{c_\ell}{c_k}\frac{R_{aj} - R_{ai} + R_{bi} - R_{bj}}{2(1 - \beta_\ell R_{ij})}\,.$$

The extension of LODFs for a joint failure follows. Denote $\mathcal{J}(\ell)$ by the collection of branches that fail jointly with ℓ as

$$\mathcal{J}(\ell) = \{k \ : \ |f_k^{(\ell)}| \geq 1\}\,.$$

The redistributions amount to calculating the matrix of PTDFs of the altered network $G^{(\ell)} = G(n, m \backslash \mathcal{J}(\ell))$, denoted by $H^{(\ell)}$. After redistribution the power flows are given by

$$\boldsymbol{f}^{(\ell)} = H^{(\ell)}\boldsymbol{p}\,.$$

A disconnected branch might cause a part of the network to separate and form an *island*. Such a branch is referred to as a *bridge* between the two network components. Bridges are crucial links in the power network as their disconnection can leave islands with insufficient generation capability and vulnerable to collapse. A branch (i, j) is a bridge if it has effective resistance $R_{ij} = 1$. Power flow calculations must be performed separately for each island.

2.2 Cascading failure review

There are two types of cascade models: *physical models* and *abstract models*. The physical models calculate the power flows for a stable system (no violations in the network constraints) and introduce a cascade initiating event. Typically, the initiating event is random, in order to represent a disturbance on the network caused by an exogenous event such as extreme weather, equipment malfunction or malicious damage.

Current physical models implement different initiating events and approximate the thermal process of line overload [4, 15, 64]. In the operation of the power network, calculating the AC power flows is standard procedure. However, the physical models in the cascade literature calculate the power flows using the DC model from Equation (2.4), and determine the redistribution using the LODFs from Equation (2.5).

On the other hand, the abstract models do not calculate power flows, but instead consider a loading-dependent approximation to cascading failure. Using the loading-dependent approximation it becomes a straightforward exercise to estimate the probabilities of large-scale blackouts, which are otherwise not tractable. Additionally, the parameters of the abstract models can be estimated from cascade data. Estimation is possible due to the reduced parameter space that neglects the network information (line ratings, susceptance, etc.).

2.2.1 Physical models

Previous research examines cascading failure from a number of perspectives. Dobson *et al.* [15] consider the state of the power grid over time. The model is composed of a short-time scale and a long-time scale. The long-time scale loosely captures the idea that grid operators will upgrade the network in response to line failures, and the short-time scale consists of cascading failure. The authors consider a physical model that uses a random initiating event. This model contains two types of line failures: *initiating random line failures* and *random failures of overloaded lines*. The first type replicates a random network disturbance, where the lines are assumed to be failing independently of each other, with probability

$$\mathbb{P}(\text{line } \ell \text{ fails}) = h^0(f_{\ell,t})\,,$$

where h^0 is a positive and non-decreasing function and $f_{j,t}$ is the power flow along line ℓ on day t. The authors choose h^0 to be a constant function, which intuitively means that failure is irrespective of the power flows. This is reasonable if we assume that the disturbance does not target specfic network elements. However, disturbances caused by strong winds are highly likely to target specific regions of the network due to the geographical dependence. For instance, transmission lines located in areas prone to strong winds are more likely to disconnect.

The second type of failure is an approximation to the heating of a line. An overloaded line takes time to heat up and the overload may be corrected prior to failure. The overloaded lines are assumed to be failing independently of each other, with probability

$$\mathbb{P}(\text{overloaded line } \ell \text{ fails}) = h^1(f_{\ell,t})\,,$$

where h^1 is a positive and non-decreasing function of the overload $|f_{\ell,t}| \geq 1$. The authors also choose h^1 to be constant. This is a crude estimate as lines with larger power flows will heat up faster, making them more likely to fail than lines that are marginally overloaded. Such behaviour is not represented in the assumption of uniform probability. However, this initial model for cascading failure defines a clear structure which other authors have since refined.

Bernstein *et al.* [4] develop a short-time cascade model targeted towards physical attacks, such as an *electromagnetic pulse* (EMP). The model introduces a random initiating event that is isotropic, such that all system elements within the radius of the epicentre are disconnected. The authors improve upon the approximation of the thermal dynamics presented in Dobson *et al.* [15] by adopting an *exponential*

smoothing process. Specifically, denote by $\tilde{f}_{\ell,t}$ the smoothed process of the power flow along line ℓ at stage t, then

$$\tilde{f}_{\ell,t} = \alpha|f_{\ell,t}|+(1-\alpha)\tilde{f}_{\ell,t-1} \text{ for } t \geq 1\,,$$

where the *smoothing factor* $\alpha > 0$ controls the lag due to the thermal dynamics, such that when $\alpha = 1$ the smoothed process is equal to the power flow. A line failure occurs when the smoothed process experiences overload such that $\left|\tilde{f}_{\ell,t}\right| \geq 1$. The smoothed process is an improvement upon the uniform dynamics presented by Dobson *et al.* [15] because lines with larger overloads fail faster than lines with smaller overloads. Additionally, this framework allows the parameter α to be tuned to achieve a deterministic time to failure. This is a more dynamic failure process which could be extended to incorporate daily ambient temperature to tune the lag parameter. The authors ran simulations of their model on the Western Electricity Coordinating Council (WECC) regional interconnect power grid in America. The authors concluded that the San Diego area was most vulnerable, which was very topical at the time due to the recent large-scale blackout in 2011.

Wang *et al.* [64] introduce a stochastic Markov chain model based on power flow redistribution which incorporates the uncertainty in demand and renewable generation. It is more realistic to treat the power injections as a stochastic input rather than a deterministic input due to the uncertainties involved in forecasting renewable generation. The Markov model is defined on the state space given by the vector of line states $\boldsymbol{s} = [s_1, \ldots, s_m]$, where $s_\ell \in \{0, 1\}$ such that $s_\ell = 0$ represents a tripped line and $s_\ell = 1$ represents a line that is still operational. Figure 2.2 depicts the line state transition process where $\lambda_\ell(t)$ is the line failure rate and $\mu_\ell(t)$ is the line restoration rate. The restoration rate is included for generality but is held fixed at $\mu_\ell(t) = 0$, as line restoration usually occurs after the cascade stops, and therefore $s_\ell = 0$ is an absorbing state. The significance of this method is that $\lambda_\ell(t)$ is a time-varying process that depends on the power flows. This allows the authors to disconnect lines based on the relay settings specificied to trip overloaded lines within a time threshold, which is potentially different for each line. Realistic relay settings are more accurate line tripping mechanisms than exponential smoothing implemented by Bernstein *et al.* [4], but the trade-off is increased model complexity.

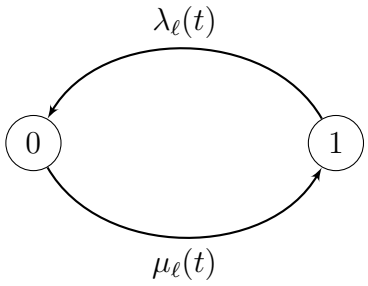


Figure 2.2: Line state transitions.

Nesti *et al.* [38] build on the idea of stochastic power injections introduced by Wang *et al.* [64] in order to analyse cascades that arise from fluctuations in renewable generation, called *emergent cascades*. Specifically, for a network $G(n, m)$ they model the power injections, $\boldsymbol{p}$, as an n-dimensional multivariate Gaussian vector such that

$$\boldsymbol{p} \sim N_n(\boldsymbol{\mu}, \varepsilon \boldsymbol{\Sigma}_p)\,,$$

where $\boldsymbol{\mu}$ represents the nominal values of the power injections, $\boldsymbol{\Sigma}_p$ represents the stochastic fluctuation of the power injections, and the parameter $\varepsilon > 0$ is used to control the magnitude of fluctuation. The linear relation between the power flows and the power injections is leveraged to estimate the distributional properties of the power flows. The authors use large deviations theory to analyse emergent failures and their propagation in power networks. In particular, they demonstrate that emergent cascades have a higher line average propagation than classical cascades, and therefore are more likely to produce fast-moving large-scale blackouts.

2.2.2 Abstract models

The CASCADE model was developed by Dobson *et al.* [16], which models cascading failure in a loading-dependent system. The authors consider a system where each line ℓ has an initial load $|\widehat{f}_\ell|$ chosen uniformly from a specified interval. Additionally, the branch capacity, c_ℓ, is included such that if at any stage during the cascade $\left|\widehat{f}_\ell\right| \geq c_\ell$ the line ℓ is disconnected. The cascade is initiated by some initial disturbance which adds the load D onto every line, large enough to trip at least one line. Power that is redistributed, due to each line failure event, places more stress on the operational lines. This is approximated by adding the load P to every operational line. Denote by M_t the number of line failures in stage t. Then the number of component failures

prior to stage t is given by

$$S_{t-1} = \sum_{j=0}^{t-1} M_j \,, \text{ for } t \geq 1, \text{ and } t \in \mathbb{N} \,.$$

Given they have not failed in a previous stage, the lines that fail at stage t of the cascade satisfy

$$|\widehat{f}_\ell| + D + S_{t-1}P \geq c_\ell \,.$$

The model is then used in Dobson *et al.* [18] to demonstrate the risk of cascading failure caused by system loading. At very low loading, failures have a minimal impact upon the network and the distribution of cascade size has an approximate exponential tail. However, when the system is loaded near a critical point, the distribution of cascade size exhibits a power law region that increases the risk of large-scale blackouts. The authors demonstrate that the distribution of the final cascade size, S, is given by

$$\mathbb{P}(S = r) \approx \theta(r\lambda + \theta)^{r-1} \frac{e^{-(r\lambda+\theta)}}{r!} \,, \text{ for } 0 \leq r \leq n-1 \,,$$

a generalised Poisson distribution with parameters θ and λ. In another paper of Dobson *et al.* [17], the same authors show that this is equivalent to the total number of individuals produced by a branching process, where the initial number of failures is Poisson with parameter θ and the number of failures in consecutive stages is Poisson with parameter λ. It is particularly useful to formulate the *Galton-Watson branching process* to statistically describe cascading failure.

Cascades are comprised of a series of stages, where the initiating event is stage 0. The Galton-Watson branching process [30] evolves from power redistribution where each line failure i in stage t produces a random number ξ_i^t of line failures in stage $(t+1)$ as described in Definition 2.2.1.

Definition 2.2.1. A *Galton-Watson* Poisson branching process $\{M_t\}_{t\in\mathbb{N}}$ evolves according to the recurrence relation $M_0 = 1$ and for $t > 0$

$$M_{t+1} = \sum_{i=1}^{M_t} \xi_i^t \,,$$

where ξ_i^t are independent and identically distributed Poisson random variables with parameter $\lambda > 0$.

A simple pictorial representation of a Galton-Watson process can be seen in Figure 2.3. However, the cascade event is actually a random number M_0 of independent Galton-Watson processes. The importance of the branching processes is that their

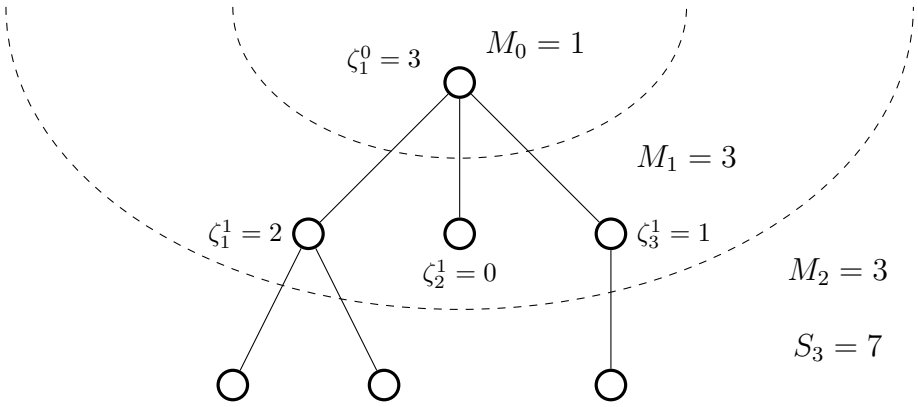


Figure 2.3: A realisation of a Galton-Watson branching process where the nodes indicate line failures and the edges indicate dependence between failures.

parameters are, as shown in Section 2.2.3, estimable and can be used to approximate the distribution of cascade size. In contrast, regular simulation methods are inefficient at obtaining statistics for large-scale blackouts.

2.2.3 Validation

Cascading failure has three main features: *line failure propagation*, *cascade size* and *spatial spreading*. The branching process approximation to cascading failure quantifies line failure propagation and predicts cascade size. Qi *et al.* [55] estimate the parameters of the branching process, θ and λ, from cascade data outlined in Table 2.1. The maximum likelihood estimator for the initial number of failures is given by

$$\hat{\theta} = \frac{1}{K} \sum_{k=1}^{K} M_0^{(k)},$$

where the superscript, $^{(k)}$, denotes cascade k of the total K observed cascades. Also, the standard Harris estimator of propagation is given by

$$\hat{\lambda} = \frac{\sum_{k=1}^{K} \left(M_1^{(k)} + M_2^{(k)} + \cdots \right)}{\sum_{k=1}^{K} \left(M_0^{(k)} + M_1^{(k)} + \cdots \right)}.$$

The estimators $\hat{\theta}$ and $\hat{\lambda}$ represent the average number of lines that jointly fail in the initial failure event and the average line failure propagation, respectively. A cascade model is validated by simulating cascades and comparing the MLEs $\hat{\theta}$ and $\hat{\lambda}$ to those obtained from observed cascade data.

	stage 0	stage 1	stage 2	$\cdots$
cascade 1	$M_0^{(1)}$	$M_1^{(1)}$	$M_2^{(1)}$	$\cdots$
cascade 2	$M_0^{(2)}$	$M_1^{(2)}$	$M_2^{(2)}$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
cascade K	$M_0^{(K)}$	$M_1^{(K)}$	$M_2^{(K)}$	$\cdots$

Table 2.1: Structure of cascade data, where $M_t^{(k)}$ is the number of failures in stage t of cascade k.

Dobson *et al.* [19] validate the spatial spreading using outage data of the form presented in Table 2.1. The outage data consists of a list of transmission line names accompanied by the time of failure (to nearest minute). Consider the smallest sub-network that contains all the line failures. First, the authors group the data into individual cascades such that a time gap of more than one hour between successive failures starts a new cascade. Then, they group outages within individual cascades into stages such that a time gap of more than one minute starts a new stage. The spatial spreading of cascading failure refers to the distance between successive line failures. The authors define the notion of the bus network distance between lines ℓ and k to be

$$d^{\text{bus}}(\ell, k) = \text{minimum number of buses in a network path joining the midpoint of } \ell \text{ to the midpoint of } k\,.$$

The mean bus distance can be estimated for any network. This provides the necessary normalisation to compare spatial spreading between simulated and real cascade data.

2.3 Congestion review

Network congestion occurs whenever a transmission line is operating outside its stability limit [11]. *Congestion management* concerns the process to alleviate congested elements and avoid line failure [32]. Managing congestion is particularly challenging in power systems because the physical flows are controlled through the power injections by power flow physics. Additionally, further restrictions such as bus voltage limits and ensuring secured $N-1$ operation must be considered [14].

Congestion management occurs over different time frames. Short-term management

consists of *generation re-dispatch*, where network operators re-dispatch power in order to alleviate congestion and secure operation. Long-term management consists of investment in transmission capacity and network expansion in order to relieve the most frequent congestions. Over the past decade, a significant amount of new power generation has been invested in wind energy around the world [65]. Without sufficient long-term planning, increased wind capacity has resulted in frequent network congestion and has necessitated frequent short-term management [29, 33].

Frequent short-term congestion management may result in a significant increase in operational costs [29, 63]. The cost of re-dispatch actions primarily includes compensation to conventional and/or renewable generators, for any additional or curtailed output. Curtailment is a reduction in the output of a generator from what it could otherwise produce, and is usually forced by the network operators [8]. Conventional generation could be entitled to compensation for additional fuel, carbon emissions and possibly extra start-ups [63]. Effective modelling of congestion management supports network operators to expand their network appropriately in order to minimise short-term congestion management. We present power network case studies that review observed network congestion and wind curtailment. Then, we describe models that have been used for network congestion and wind curtailment studies.

2.3.1 Case studies

High wind penetration levels induce congestion into the network due to two main factors:

1. negative temporal correlation between wind power and demand, and
2. geographical mismatch between wind power and demand in the network [24].

Negative temporal correlation occurs because it is generally windier at night when demand is low, and calmer in the day when demand is high. Geographical mismatch refers to the fact that wind farms are typically built in rural areas, as they require large amounts of land space, whilst demand is centered in highly-populated cities. This results in higher power transfers through rural areas of the network that are not necessarily designed to support such load. Consequently, wind generators are curtailed to alleviate congestion in the network [24, 29].

Joos and Staffell [29] investigate the effects of wind curtailment in Germany and Britain over the period of 2012 to 2016. The curtailment rate is the measure of

the percentage of time wind curtailment is present. Higher curtailment rates cause an increase in the total curtailed wind output, which then directly impacts the compensation paid to wind farms for curtailment. Table 2.2 highlights key statistics relevant to quantifying wind curtailment.

Germany	2012	2016
Wind generation (%)	8.37	13.02
Wind curtailment rate (%)	0.70	4.36
Curtailed wind output (TWh)	0.36	3.53
Compensation to wind (m€)	30.85	325.89
Britain		
Wind generation (%)	5.61	11.13
Wind curtailment rate (%)	0.44	5.64
Curtailed wind output (TWh)	0.05	1.12
Compensation to wind (m€)	5.92	81.88

Table 2.2: Summary statistics of wind curtailment in Germany and Britain.

From Table 2.2, it is evident that higher wind penetration levels induce wind curtailment. For what is a fairly modest increase in percentage of wind generation, both countries experience significantly larger increases in curtailed wind output. This means that without effective congestion management further investment in wind power will be less economical.

Germany's large investment in wind power has resulted in higher wind penetration but at a significant cost. Further investment in wind power is likely to aggravate congestion. In Britain, the authors identify that a large portion of the onshore wind farms located in Scotland (North of Britain) demonstrate high curtailment rates. The main contributing factor is geographical mismatch, with wind power centred in Scotland whilst demand is centred in England and a weak North-South connection in the power network. The operators have undertaken a long term strategy to reinforce the North-South network to alleviate congestion.

Similarly, Luo *et al.* [33] determine that wind curtailment in China is caused by a combination of geographical mismatch and weak network transmission to central China. The distribution of effective wind power in China is centered in the northern regions with no useable wind speeds in central China [59]. From 2000 to 2012, installed wind capacity in China increased by 48% annually. However, the authors

highlight that proper network planning was not undertaken, resulting in weak network connections to central China regions and increased network congestion.

2.3.2 Re-dispatch and curtailment models

Power is re-dispatched in order to alleviate congestion within a short-term time scale. This secures operation and avoids the potential of line failure. However, as a result of increased security constraints, re-dispatch may curtail wind generation. Typically, re-dispatch models are a variation on a DC optimal power flow.

Gu *et al.* [24] quantify the impact of congestion on curtailment by defining the sensitivity of wind curtailment for line ℓ as follows

$$s_\ell = \frac{\partial W^C}{\partial c_\ell},$$

where W^C is the amount of wind curtailment and c_ℓ is the line capacity. For a power network with one congested line, ℓ, and one curtailed wind farm at bus, k, a perturbation applied to the capacity of line, ℓ, results in the following system of differential equations

$$\partial p_k + \partial p_h = 0, \tag{2.8}$$

$$H_{\ell k}\partial p_k + H_{\ell h}\partial p_h = \partial c_\ell. \tag{2.9}$$

Here, p_k and p_h are the power injections at bus k and h, respectively, H is the matrix of PTDFs, (2.8) is the power balancing equation where the marginal unit at bus h picks up the power mismatch due to curtailment, and (2.9) is a result of the DC model from Equation (2.4). The analytical expression of wind curtailment sensitivity is given by

$$s_\ell = \frac{-\partial p_k}{-H_{\ell h}\partial p_k + H_{\ell k}\partial p_k} = \frac{1}{H_{\ell k} - H_{\ell h}},$$

and indicates the curtailment to be relieved by increasing the capacity, c_ℓ. The authors perform a simulation using a DC optimal power flow such that the curtailed generator and marginal unit are chosen to minimise costs. The same authors extend their initial model to allow for multiple congested lines [23]. However, this technique targets one wind farm and curtails its output, as opposed to managing the output of all generators, and therefore may over-estimate wind curtailment that occurs in a real power network.

More general optimal power flows that re-dispatch all generators have been used in other curtailment studies. Burke and O'Malley [12] use a DC security-constrained

optimal power flow in order to perform a curtailment sensitivity analysis. This is a linear optimisation model that applies $N-1$ contingency as the security constraint. In addition to the security constraint, McKenna *et al.* [35] incorporate voltage control, non-synchronous penetration limits and operating power reserves. Limits on non-synchronous generation may be enforced by network operators in order to maintain required stability. Van den Bergh *et al.* [63] combine a comprehensive re-dispatch model with a day-ahead unit-commitment model. The day-ahead generation dispatch minimises operational generation costs and does not impose transmission limits. Therefore, the model of Van de Bergh *et al.* [63] estimates curtailment from the most economical generation dispatch.

2.4 Statistical concepts

This section details the fundamental mathematical methods used in this thesis. Initially, we introduce the theory of *large deviations* in order to evaluate the probability of rare events. Then, we introduce *time series*, and explain different modelling techniques such as *Fourier series*, *auto-regressive moving average (ARMA) processes*, and relevant goodness-of-fit diagnostic plots. Finally, we introduce *least squares regression* and its extension to a generalised non-parametric *quantile regression*.

2.4.1 Large deviations

Consider the *mean* of a sequence of n real random variables $X_1, X_2, \ldots, X_n$ given by

$$S_n = \frac{1}{n}\sum_{i=1}^{n} X_i \,.$$

The *Weak Law of Large Numbers* states that if $X_1, X_2, \ldots, X_n$ are independent and identically distributed with $\mu = \mathbb{E}[X_1] < \infty$ then for every $\varepsilon > 0$

$$\mathbb{P}(|S_n - \mu| > \varepsilon) \to 0\,, \text{ as } n \to \infty\,.$$

Whilst this provides understanding of the mean in the limit as $n \to \infty$, it does not describe the rate of convergence. The theory of large deviations shows that in many cases the *decay rate* is in fact exponential. Consider the probability that the mean S_n exceeds $a > \mu$. Then, the *Chernoff bound* is an upper bound given by

$$\mathbb{P}\left(S_n > a\right) \leq e^{-n\delta(a)}\,, \tag{2.10}$$

where $\delta(a)$ is known as the *decay rate* [61]. Without loss of generality, the decay rate is a *Legendre transform* of the random variable X_1 such that

$$\delta(a) = \sup_{\theta}(\theta a - \log M(\theta)) \,, \tag{2.11}$$

where $M(\theta) < \infty$ is the moment generating function of X_1. A direct result from Equation (2.10) is *Cramer's theorem* given by

$$\lim_{n\to\infty} \frac{1}{n} \ln \mathbb{P}(S_n > a) = -\delta(a) \,.$$

2.4.2 Time series

We provide a brief introduction to time series analysis, and for further detail see [25]. A *time series* is an ordered sequence of observations recorded over time, where each observation is time dependent. Generally, we analyse discrete-time series where the observations y_t are recorded at specific times even when the measured variable Y_t is continuous. Discrete-time series contain values of the measured variable at equally-spaced times, in order to create either a *sampled series* or an *averaged series*. A sampled series records the observed value at a specific time, whereas an averaged series records the average of the observed values over a specific interval. Due to time-dependent observations, time series models explain the temporal correlation in the observed data. Models that explain the temporal dependence can predict future values based on previous observations. Processes that can be predicted exactly are known to be deterministic, whereas stochastic processes incorporate random behaviour such that future values have a probability distribution conditioned on past values.

The simplest stochastic process is a random process where each observation is independent, called the *white noise process*.

Definition 2.4.1. A process $\{\varepsilon_t\}_{t\in\mathbb{N}}$ is said to be *white noise* with mean 0 and variance σ^2, written

$$\varepsilon_t \sim N(0, \sigma^2)$$

if the random variables ε_t are independent and identically distributed normal random variables.

By definition, it is immediate that the first two moments, mean and variance, are constant over time. Since the observations are identically distributed, it follows that higher-order moments also remain constant over time. Such a process is referred to as being *strictly stationary*.

Definition 2.4.2. A process is *strictly stationary* if

$$(Y_{t_1}, \ldots, Y_{t_k}) \text{ and } (Y_{t_1+\tau}, \ldots, Y_{t_k+\tau})$$

have the same joint distribution for all sets of time points $t_1, \ldots, t_k$ and all integers $\tau \geq 0$, for $k \geq 1$.

Very few processes are strictly stationary. In practice we are concerned only with *second-order stationary*, where only the first two moments remain constant over time.

Definition 2.4.3. The process $\{Y_t\}_{t\in\mathbb{N}}$ is called *second-order stationary* if

$$\begin{aligned} \mathbb{E}(Y_t) &= \mu\,, \\ cov(Y_t, Y_{t-\tau}) &= \gamma(\tau)\,, \end{aligned}$$

for all $\tau \in \mathbb{N}$ and the auto-covariance function $\gamma(\tau)$ depends only on the lag τ.

The definition of second-order stationary involves the *auto-covariance function*, $\gamma(\tau)$, which describes the second-order variation between observations in time. The auto-covariance function is normalised to values between 1 and -1 to obtain the auto-correlation function. Auto-correlation is a measure of the dependence between values observed at different lags, τ, and is a useful diagnostic tool to assess whether a model captures the temporal dependence within the data.

Definition 2.4.4. The process $\{Y_t\}_{t\in\mathbb{N}}$ with auto-covariance function $\gamma(\tau)$ has *auto-correlation* function given by

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)} \text{ for } \tau \geq 0\,.$$

At lag $\tau = 0$, the auto-correlation is always perfectly correlated such that $\gamma(0) = 1$, and at lag $\tau = 1$, the auto-correlation is measured using adjacent observations. The auto-covariance and auto-correlation is generalised to the *cross-covariance* and *cross-correlation* in order to observe the relation between two time series.

Definition 2.4.5. The processes $\{X_t\}_{t\in\mathbb{N}}$ and $\{Y_t\}_{t\in\mathbb{N}}$ have *cross-covariance* at lag τ given by

$$\gamma_{xy}(\tau) = \frac{1}{n-1}\sum_{t=1}^{n}(X_{t-\tau} - \mu_x)(Y_t - \mu_y)\,,$$

where μ_x and μ_y are the means of the processes $\{X_t\}_{t\in\mathbb{N}}$ and $\{Y_t\}_{t\in\mathbb{N}}$, respectively, and n is the number of samples.

Definition 2.4.6. The processes $\{X_t\}_{t\in\mathbb{N}}$ and $\{Y_t\}_{t\in\mathbb{N}}$ with *cross-covariance* function $\gamma_{xy}(\tau)$ have *cross-correlation* at lag τ given by

$$\rho_{xy}(\tau) = \frac{\gamma_{xy}(\tau)}{\sqrt{\gamma_{xx}(0)\gamma_{yy}(0)}}\,,$$

where $\gamma_{xx}(0)$ and $\gamma_{yy}(0)$ are the variances of the processes $\{X_t\}_{t\in\mathbb{N}}$ and $\{Y_t\}_{t\in\mathbb{N}}$, respectively.

In comparison to auto-covariance, the cross-covariance, and therefore the cross-correlation, is an asymmetric function such that $\rho_{xy}(\tau) \neq \rho_{xy}(-\tau)$. A peak in the cross-correlation at lag τ indicates that one time series experiences a delay τ.

It is common to decompose a time series into the *deterministic* and *stochastic* components. Consider the decomposition of the time series $\{Y_t\}_{t\in\mathbb{N}}$ such that

$$Y_t = D_t + \psi_t\,,$$

where D_t represents the deterministic component and ψ_t represents the stochastic component. The deterministic component contains the overall trend and seasonality, whilst the stochastic component is a stationary series of random variation. *Fourier series* is a spectral analysis technique used to identify significant cycles that can be included in the deterministic component [9]. Alternatively, ARMA processes are probabilistic models extensively used for modelling stationary time series.

2.4.3 Fourier series

Fourier transforms provide a method for frequency domain analysis or spectral analysis. In the frequency domain, seasonal cycles are more easily detected than in the time domain. The Fourier transform of a function $f(t)$ is given by

$$F(\omega) = \int_{-\infty}^{\infty} f(t)\exp^{-2\pi i\omega t}\,dt\,,$$

where t is time and ω is frequency. Additionally, the original function $f(t)$ can be expressed as

$$f(t) = \int_{-\infty}^{\infty} F(\omega)\exp^{2\pi i\omega t}\,dt\,,$$

which is known as the *Fourier inversion theorem* [22].

As time series data consists of discrete observations of a continuous variable, we focus on the discrete Fourier transform. Consider the decomposition of the time

series $\{y_t : t = 1, 2, \ldots, n\}$ into trigonometric functions such that

$$y_t = \mu + \sum_{j=1}^{\frac{n}{2}-1} \left(\alpha_j \cos\left(\frac{2\pi jt}{n}\right) + \beta_j \sin\left(\frac{2\pi jt}{n}\right) \right) ,$$

known as a *Fourier series*. The Fourier series demonstrates that the summation of cosine and sine waves at different frequencies can represent any series. Each frequency contributes a certain amount to the variance of the series, called the *power*. A frequency with a high power is more significant than another frequency with a lower power. The distribution of power across the frequencies is called the *power spectrum*, defined from the least squares estimates of the Fourier coefficients given by

$$\begin{aligned} \widehat{\mu} &= \frac{1}{n} \sum_{t=1}^{n} y_t \,, \\ \widehat{\alpha}_j &= \frac{2}{n} \sum_{t=1}^{n} y_t \cos\left(\frac{2\pi jt}{n}\right) , \\ \widehat{\beta}_j &= \frac{2}{n} \sum_{t=1}^{n} y_t \sin\left(\frac{2\pi jt}{n}\right) . \end{aligned}$$

The power spectrum is the Fourier transform of the auto-covariance function and shows how the variance of the stochastic process is distributed with frequency. The sample variance is given by

$$V = \frac{1}{n} \sum_{t=1}^{n} (y_t - \overline{y})^2 = \sum_{j=1}^{n/2-1} (\widehat{\alpha}_j^2 + \widehat{\beta}_j^2) \,,$$

where $\overline{y}$ is the sample mean of the time series $\{y_t : t = 1, 2, \ldots, n\}$ [9]. The contribution of frequency ω to the sample variance is the power spectrum defined as follows:

$$\upsilon(\omega) = \frac{\widehat{\alpha}_\omega^2 + \widehat{\beta}_\omega^2}{2} \,.$$

Significant spikes in the power spectrum correspond to cycles that contribute significantly to the variance of the time series.

2.4.4 Auto-regressive moving average processes

The moving average (MA) model is a basic development of the white noise process. The model is constructed by linear combinations of lagged elements of a purely random process.

Definition 2.4.7. A process $\{Y_t\}_{t\in\mathbb{N}}$ is said to be a *moving average* of order q if

$$Y_t = \sum_{j=0}^{q} m_j \varepsilon_{t-j} \,,$$

where m_j are the moving average coefficients, and $\{\varepsilon_t\}_{t\in\mathbb{N}}$ is the white noise process, and $q \geq 0$ is an integer.

The auto-regressive (AR) model, as the name suggests, regresses future values on past values.

Definition 2.4.8. A process $\{Y_t\}_{t\in\mathbb{N}}$ is said to be *auto-regressive* of order p if

$$Y_t = \varepsilon_t + \sum_{k=1}^{p} a_k Y_{t-k} \,,$$

where a_k are the auto-regressive coefficients, $\{\varepsilon_t\}_{t\in\mathbb{N}}$ is the white noise process, and $p \geq 1$ is an integer.

The AR and MA models can be combined to form an auto-regressive moving average (ARMA) model.

Definition 2.4.9. A process $\{Y_t\}_{t\in\mathbb{N}}$ is said to be an ARMA(p, q) process with p auto-regressive and q moving-average components if

$$Y_t = \sum_{k=1}^{p} a_k Y_{t-k} + \sum_{j=0}^{q} m_j \varepsilon_{t-j} \,,$$

where a_k and m_j are the auto-regressive and moving average coefficients, respectively, $\{\varepsilon_t\}_{t\in\mathbb{N}}$ is the white noise process, and $p \geq 1$, $q \geq 0$ are integers.

Additionally, there exists a variant of ARMA models where the assumption of the error term is relaxed to be uncorrelated but dependent. In this case the variance is conditional, and forms the basis for *auto-regressive conditional heteroscedastic* (ARCH) and *generalised auto-regressive conditional heteroscedastic* (GARCH) processes, used in wind and solar time series analysis [10].

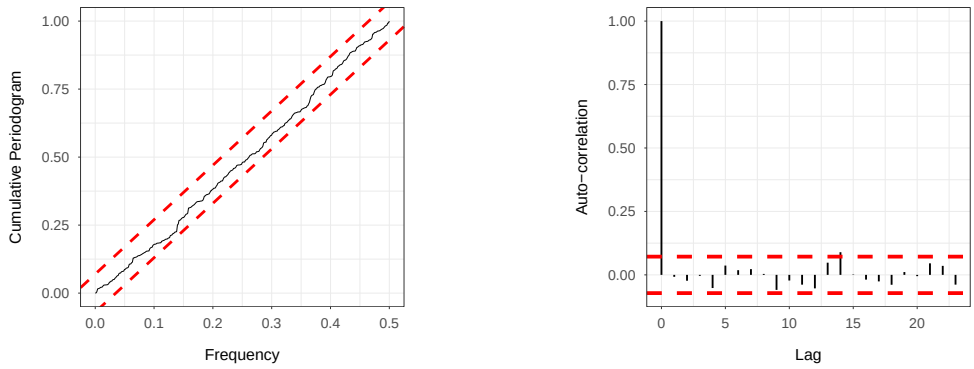
2.4.5 Goodness-of-fit

The goodness-of-fit of a statistical model describes how well the model fits the data. Diagnostic plots are used to analyse the goodness-of-fit to observe patterns in the residuals. Identifying patterns in the residual behaviour may inform a more accurate model. We perform two diagnostic plots: *cumulative periodogram* and *auto-correlation*. The cumulative periodogram is, as the name suggests, the cumulative

values of the periodogram. The periodogram is a representation of the spectral density similar to the power spectrum described in Section 2.4.3. The residuals of a model with a reasonable goodness-of-fit will appear as white-noise in the diagnostic plots.

Recall from Definition 2.4.1, that the white noise process consists of independent observations. This implies there is no correlation between any two observations. Therefore, the periodogram is flat because there are no statistically significant cycles, and the auto-correlation is effectively zero for all lags greater than zero. At lag 0, the time series is perfectly correlated with itself, which is a feature of any time series.

We simulate a white-noise process and present the diagnostic plots in Figure 2.4. We include the hypothesis test for white noise at the 5% level of significance, indicated by the red lines. The null hypothesis is that the residuals are white noise. The null is rejected when values violate the 5% significance limits. In particular, a spike in the cumulative periodogram (Figure 2.4a) indicates a significant cycle at the corresponding frequency and a spike in the auto-correlation (Figure 2.4b) indicates a significant auto-correlation at the corresponding lag. Both of which indicate deviations from white noise.



(a) Cumulative periodogram.

(b) Auto-correlation.

Figure 2.4: Goodness-of-fit diagnostic plots of a white noise process. The dashed red lines indicate the 5% significance levels of the corresponding hypothesis test. (a) The cumulative periodogram where the x-axis represents the frequency of the cycle. (b) The auto-correlation where the x-axis is the lag. From the diagnostics we would conclude that the process is white noise.

The cumulative periodogram in Figure 2.4a is linear and remains within the 5% level of significance. The auto-correlation in Figure 2.4b is not statistically significant for

lags greater than zero. Since there are no violations the diagnostics demonstrate a process consistent with white noise. However, we highlight that even the white noise process may violate the significance level. Therefore, the hypothesis test is not a hard cut-off and may require additional analysis.

2.4.6 Linear regression

Linear regression is a statistical method used to describe the mean effects of a set of predictor variables on a response variable. A simple linear regression relates the *response variable* Y and *predictor variable* X as follows:

$$Y = \beta_0 + \beta_1 X + \varepsilon \,,$$

where β_0 and β_1 are unknown parameters, and ε is the error due to a number of possible factors, including the choice of the model [62]. Given the observed set of data pairs (x_i, y_i) for $i = 1, \ldots, n$ of the random variables X and Y, the unknown parameters β_0 and β_1 are estimated using the *least squares method*. The *least squares estimates* $\widehat{\beta}_0$ and $\widehat{\beta}_1$ minimise the sum of squared errors given by the quantity

$$\sum_{i=1}^{n} \left[y_i - (\beta_0 + \beta_1 x_i) \right]^2 .$$

It is convenient to represent the regression model in matrix notation such that

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} ,$$

which simplifies to

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \,,$$

where $\boldsymbol{X}$ is often referred to as the model matrix. From linear algebra, the least squares estimates $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are given by

$$\widehat{\boldsymbol{\beta}} = \begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \end{bmatrix} = \left(\boldsymbol{X}^\top \boldsymbol{X} \right)^{-1} \boldsymbol{X}\boldsymbol{Y} \,.$$

The *fitted* and *residual* values are given by $\widehat{\boldsymbol{Y}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\varepsilon}} = \boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}$, respectively. Additional predictor variables may be included in the model for multiple regression analysis. The key assumptions for regression are:

1. **The relationship between the response variable and each of the predictor variables is linear.** However, the predictor variables may be transformed in order to obtain a linear relationship.

2. **The residual values are homoscedastic.** The residuals $\widehat{\boldsymbol{\varepsilon}}$ are assumed to have constant variance.

After ensuring the model assumptions are satisfied and the model is unbiased, the next step of model evaluation is measure of *goodness of fit*. Numerical quantities such as *R-squared* (R^2), or *coefficient of determination*, are used to evaluate the residual spread around the fitted regression line. R-squared is a measure of the explained sum of squares over the total sum of squares such that

$$R^2 = 1 - \frac{\sum_i (y_i - \widehat{y}_i)^2}{\sum_i (y_i - \overline{y})^2} \,,$$

where $\overline{y}$ is the sample mean of y_i, $i = 1, \ldots, n$. For the same dataset, higher R-squared values indicate a smaller residual spread. As predictor variables are added to the model, the R-squared value will always increase; therefore, a model with more predictor variables will appear to have a better fit. In multiple regression, the *adjusted R-squared* is used because the quantity adjusts for the number of predictors in the model, and increases only if a new predictor improves the model fit more than expected. The adjusted R-squared value is given by

$$\overline{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-p-1} \,,$$

where p is the total number of predictor variables and n is the sample size.

2.4.7 Quantile regression

For a response Y and predictor X, linear regression models the conditional mean $\mathbb{E}[Y|X]$ and assumes the conditional distribution is normal with constant variance. On the other hand, quantile regression extends linear regression to model different quantile levels of the conditional distribution, refer to [31] for more details.

The quantile level, τ, is the proportion associated with the corresponding conditional quantile of Y given X, written as $Q_{Y|X}(\tau)$. Therefore, the quantile level is the value of Y below which the proportion of the conditional response is τ such that

$$\tau = \mathbb{P}\left(Y \leq Q_{Y|X}(\tau)|X\right) \,.$$

The quantile regression model assumes that the τth conditional quantile function on p predictor variables is given by

$$Q_{Y|X}(\tau) = \boldsymbol{X}\boldsymbol{\beta}_\tau \,,$$

where $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ is the model matrix, and $\boldsymbol{\beta}_\tau \in \mathbb{R}^{p \times 1}$ is the column vector of quantile regression estimates. The quantile regression estimates of $\boldsymbol{\beta}_\tau$ are given by

$$\widehat{\boldsymbol{\beta}}_\tau = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \rho_\tau \left(y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta} \right) ,$$

where $\boldsymbol{x}_i$ is the column vector of the ith row of the model matrix $\boldsymbol{X}$, and ρ_τ is the *check function* given by

$$\rho_\tau(u) = u(\tau - \mathbb{1}_{u<0}) .$$

Chapter 3

Emergent cascades

The majority of the literature analyses cascades that arise from a random physical disturbance [6, 15, 64]; these are referred to as *classical cascades*. In recent years, power networks have transitioned to use more renewable resources and less conventional power. Renewable resources such as wind and solar are known to be stochastic, as they depend on the available resource. Whilst some models consider power generated from renewable sources to be stochastic [64], Nesti *et al.* [38] analyse their potential to cause the initial line failure. This is referred to as an *emergent failure* and the redistribution of power may result in an *emergent cascade*.

First, we present the *emergent cascade* model of Nesti *et al.* [38], which includes their theoretical model as well as a case study of the German power network. In Section 3.2, we analyse the goodness-of-fit of the models for wind and solar generation used by Nesti *et al.* [38], and motivate our extension to the model of solar generation. Then, in Section 3.3, we extend the model of solar generation and demonstrate its improved goodness-of-fit. We conclude this section with a discussion of our contribution and future work.

3.1 Emergent cascade model

3.1.1 Stochastic power injections

Nesti *et al.* [38] model the power injections, $\boldsymbol{p}$, as an n-dimensional multivariate Gaussian vector such that

$$\boldsymbol{p} \sim N_n(\boldsymbol{\mu}, \varepsilon\Sigma_p)\,,$$

where $\boldsymbol{\mu}$ represents the nominal values of the power injections, Σ_p represents the stochastic fluctuation of the power injections, and $\varepsilon > 0$ is used to control the magnitude of fluctuation. The distribution of the power injections, $\boldsymbol{p}$, is valid only when the covariance matrix, Σ_p, is positive definite. In order to ensure this holds, the power flow calculation is decomposed into its deterministic and stochastic parts. This is a necessary decomposition, as the deterministic part does not vary and thus corresponds to a zero diagonal entry in the covariance matrix, Σ_p, which violates positive definiteness. The deterministic part consists of conventional power and load, whereas the stochastic part consists of wind and solar power. Therefore, a stochastic bus is defined to be any bus where a wind farm and/or a solar farm is attached.

The authors assume DC power flows $\boldsymbol{f}$ is such that $\boldsymbol{f} = H\boldsymbol{p}$, where H is the matrix of PTDFs from Equation (2.4). Then, the decomposition is performed as follows:

n_s	number of stochastic buses,
n_d	number of deterministic buses,
m	number of lines,
$\mathcal{I}_s \subset \{1, \ldots, n_s\}$	indices of stochastic buses,
$\mathcal{I}_d \subset \{1, \ldots, n_d\}$	indices of deterministic buses,
$\boldsymbol{p}_s = (p_i)_{i \in \mathcal{I}_s} \in \mathbb{R}^{n_s}$	stochastic power injection,
$\boldsymbol{p}_d = (p_i)_{i \in \mathcal{I}_d} \in \mathbb{R}^{n_d}$	deterministic power injection,
$H_s \in \mathbb{R}^{m \times n_s}$	matrix consisting of the columns of H indexed by I_s,
$H_d \in \mathbb{R}^{m \times n_s}$	matrix consisting of the columns of H indexed by I_d,
$\boldsymbol{f}_s = H_s\boldsymbol{p}_s \in \mathbb{R}^m$	stochastic component of $\boldsymbol{f}$,
$\boldsymbol{f}_d = H_d\boldsymbol{p}_d \in \mathbb{R}^m$	deterministic component of $\boldsymbol{f}$.

The decomposition of the normalised power flows results in $\boldsymbol{f} = \boldsymbol{f}_s + \boldsymbol{f}_d$.

The stochastic power injections are modelled by an n_s-dimensional multivariate Gaussian with mean $\boldsymbol{\mu}_s \in \mathbb{R}^{n_s}$ and covariance matrix $\Sigma_{p_s} \in \mathbb{R}^{n_s \times n_s}$ such that

$$\boldsymbol{p}_s \sim N_{n_s}(\boldsymbol{\mu}_s, \varepsilon\Sigma_{p_s}),$$

where the $n_s \times n_s$ stochastic covariance matrix, Σ_{p_s}, is positive definite. From linear combinations of multivariate Gaussian distributions, the stochastic line covariance is given by

$$\Sigma_{f_s} = H_s\Sigma_{p_s}H_s^\top.$$

Typically, the number of transmission lines exceeds the number of buses, $m > n \geq n_s$, thus Σ_{f_s} is not positive definite and the joint distribution is degenerate. Therefore, whilst the decomposition provides insight into the stochastic and deterministic power flows, it does not provide a more complicated multivariate distribution. Hence, the marginal distributions of the line power flows are used to determine the probability of line failures.

3.1.2 Probability of line failures

Nesti *et al.* [38] assume that line failure occurs when the absolute normalised power flow along line ℓ, $|f_\ell|$, is greater than or equal to 1. However, line failure is not immediate, as it takes time for lines to heat up, and therefore is a questionable assumption. The absolute line power flows for each line ℓ follow a Gaussian distribution given by

$$|f_\ell| \sim N(|\nu_\ell|, \varepsilon\sigma_\ell^2)\,,$$

where $\nu_\ell = \boldsymbol{e}_\ell^\top(H_s\boldsymbol{\mu}_s + \boldsymbol{f}_d)$ and $\sigma_\ell^2 = (\Sigma_{f_s})_{\ell\ell}$, and $\boldsymbol{e}_\ell$ is the column vector with 1 in the ℓth position and 0 elsewhere. As $\varepsilon \to 0$, line failures become rare events and the principle of large deviations applies. The marginal distribution of the power flows follow a Gaussian distribution with a corresponding Legndre transform given by

$$\delta(a) = \frac{(a - |\nu_\ell|)^2}{2\sigma_\ell^2}\,.$$

The sequence of line power flows satisfy the large deviations principle

$$\lim_{\varepsilon\to 0} \varepsilon \ln \mathbb{P}(|f_\ell| \geq 1) = -\frac{(1 - |\nu_\ell|)^2}{2\sigma_\ell^2} = -\delta_\ell\,,$$

where δ_ℓ is the decay rate of line ℓ. The most vulnerable lines have smaller decay rates and higher probabilities of failure, and the first line to fail has the smallest decay rate.

3.1.3 Most likely power injections

The authors of [38] calculate the most likely power injections, $\boldsymbol{p}^{(\ell)}$, that cause the failure of line ℓ. The stochastic power injections, $\boldsymbol{p}_s$, and absolute power flow along line ℓ, $|f_\ell|$, are jointly multivariate Gaussian. Therefore, maximising the conditional distribution is equivalent to the constrained quadratic program

$$\begin{aligned} \min_{\boldsymbol{p}_s} \quad & \tfrac{1}{2}(\boldsymbol{p}_s - \boldsymbol{\mu}_s)^\top \Sigma_{p_s}^{-1}(\boldsymbol{p}_s - \boldsymbol{\mu}_s) \\ \text{s.t.} \quad & |\boldsymbol{e}_\ell^\top(H_s\boldsymbol{p}_s + \boldsymbol{f}_d)| \geq 1\,, \end{aligned} \tag{3.1}$$

which consists of two disjoint inequality constraints, resulting in two separate optimisation sub-problems. We solve each optimisation sub-problem using the method of Lagrange multipliers. The Lagrangian of the optimisation sub-problem corresponding to the nominal power flow $\nu_\ell > 0$ is given by

$$Ł(\boldsymbol{p}_s, \lambda) = \frac{1}{2}(\boldsymbol{p}_s - \boldsymbol{\mu}_s)^\top \Sigma_{p_s}^{-1}(\boldsymbol{p}_s - \boldsymbol{\mu}_s) + \lambda(\boldsymbol{e}_\ell^\top(H_s\boldsymbol{p}_s + \boldsymbol{f}_d) - 1)\,,$$

where $\lambda > 0$ is the Lagrange multiplier. The constraint in (3.1) becomes tight as $\varepsilon \to 0$, such that $\boldsymbol{e}_\ell^\top(H_s\boldsymbol{p}_s + \boldsymbol{f}_d) = 1$, which corresponds to the failure of line ℓ. Differentiating (3.1) with respect to the stochastic power injections, $\boldsymbol{p}_s$, results in the following:

$$\frac{\partial}{\partial \boldsymbol{p}_s} L(\boldsymbol{p}_s, \lambda) = (\boldsymbol{p}_s - \boldsymbol{\mu}_s)^\top \Sigma_{p_s}^{-1} - \lambda \boldsymbol{e}_\ell^\top H_s = 0\,,$$

which implies

$$\boldsymbol{p}_s = \boldsymbol{\mu}_s + \lambda \Sigma_{p_s} H_s^\top \boldsymbol{e}_\ell\,. \tag{3.2}$$

Then, pre-multiplying Equation (3.2) by $\boldsymbol{e}_\ell^\top H_s$ results in

$$\begin{aligned} \boldsymbol{e}_\ell^\top H_s(\boldsymbol{p}_s - \boldsymbol{\mu}_s) &= \lambda \boldsymbol{e}_\ell^\top H_s \Sigma_{p_s} H_s^\top \boldsymbol{e}_\ell\,, \\ 1 - \nu_\ell &= \lambda \sigma_\ell^2\,, \text{ where } \sigma_\ell^2 = (\Sigma_{f_s})_{\ell\ell} = \boldsymbol{e}_\ell^\top H_s \Sigma_{p_s} H_s^\top \boldsymbol{e}_\ell\,, \\ \lambda &= \frac{1 - \nu_\ell}{\sigma_\ell^2}\,. \end{aligned} \tag{3.3}$$

Substituting Equation (3.3) into (3.2), and considering the optimisation sub-problem when $\nu_\ell \leq 0$, the most likely power injections that lead to the failure of line ℓ are given by

$$\boldsymbol{p}_s^{(\ell)} = \boldsymbol{\mu}_s + \frac{(\operatorname{sign}(\nu_\ell) - \nu_\ell)}{2\sigma_\ell^2} \Sigma_{p_s} H_s^\top \boldsymbol{e}_\ell\,.$$

The most likely power injections that cause the failure of line ℓ are used to calculate the most likely power flows after redistribution

$$\boldsymbol{f}^{(\ell)} = H_s \boldsymbol{p}_s^{(\ell)} + \boldsymbol{f}_d\,.$$

After redistribution, if any other lines become overloaded, they too are disconnected. This process continues to form the emergent cascade. The authors perform a case study of the theoretical emergent cascade model on the German power network.

3.1.4 Case study: German power network

The German power network is interconnected with the rest of the synchronous grid of Continental Europe that services over 500 million people in 26 countries [57].

The transmission system operators for each regional jurisdiction formed the European Network of Transmission System Operators for Electricity (ENTSO-E) in 2009.

The German Federal Ministry of Education and Research funded the project SciGRID (Scientific GRID), which developed open-source methods for the automated generation of electricity networks. The SciGRID German power network is implemented in the Power System Analysis module for Python (PyPSA). In PyPSA, the authors calculate the Optimal Power Flow (OPF) across the German power network for each hour of the day 01/01/2011.

The German power network in Figure 3.1, consists of 585 buses, 852 lines, 1423 generators, 96 transformers and 38 pump storage units. The generating units include 489 solar farms, 488 onshore wind farms, 5 offshore wind farms and the remaining 441 conventional generators. The 96 transformers are located at the buses and inter-connect two different voltage levels: 220 kV and 380 kV.

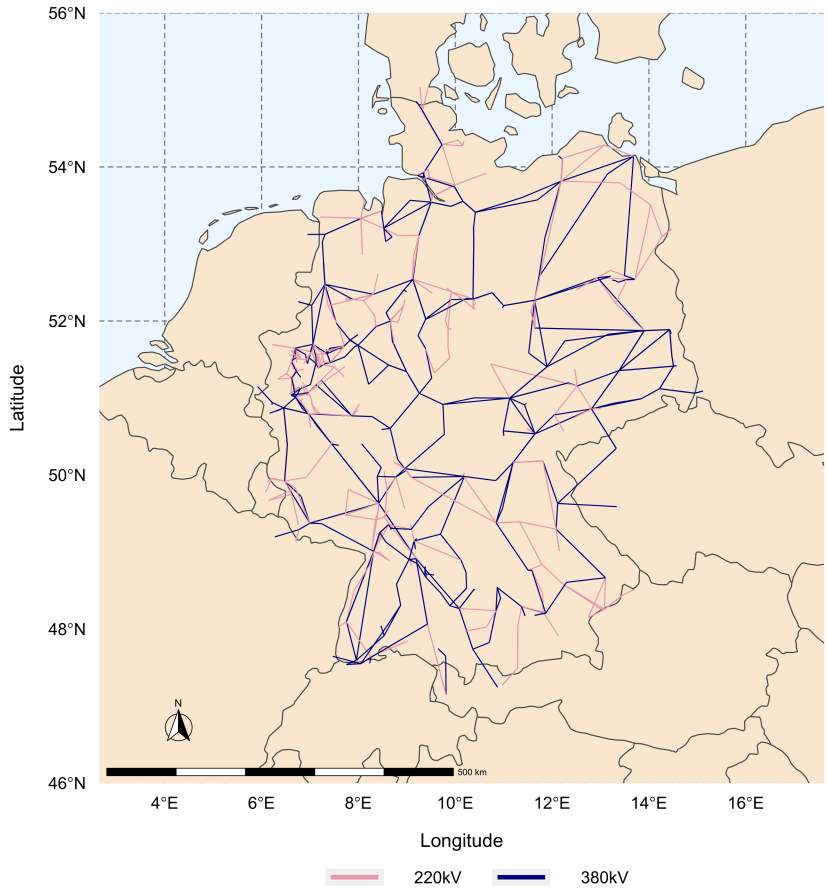


Figure 3.1: Map of the SciGRID German power network. The edge colour indicates the transmission line voltage on two levels: 220 kV and 380 kV.

The dataset includes the marginal cost of power for each fuel type. These costs are

used in the OPF in Section 3.1.5. The marginal cost of renewable resources is zero, hence multiple optimal solutions exist[1]. Also included is the hourly time series of normalised power generation (pu) of wind and solar for the year 2011. The authors exclude the storage units from the OPF in Section 3.1.5.

3.1.5 Data based model for μ_s

Nesti *et al.* [38] obtain realistic nominal power flows, $\boldsymbol{\nu}$, by performing a DC optimal power flow (OPF) for a specific hour. The OPF minimises the total cost of generation subject to energy balance, generation capacities and transmission constraints. The authors scale the line limits, c_ℓ, by a contingency factor of 0.7, in order to ensure the network is not overloaded and allow for reactive power flows. Denote by g_k the expected generation at bus k, as outputted by the OPF, and d_k the load at bus k. Then, the nominal power injections, $\boldsymbol{\mu}$, are given by

$$\mu_k = g_k + d_k \quad \text{for } k = 1, \dots, 585\,,$$

and the nominal values of the stochastic power injections are given by $\boldsymbol{\mu}_s = (\mu_k)_{k \in \mathcal{I}_s}$.

3.1.6 Data based model for Σ_{p_s}

Nesti *et al.* [38] estimate the steady-state covariance, Σ_p, from the residual fluctuation around the fitted values of a stochastic model. The authors fit ARMA models to hourly power generation values of wind and solar.

The SciGRID dataset contains the time series

$$\boldsymbol{y}_{\text{w.off}} \in \mathbb{R}^{8760\times 5}\,, \boldsymbol{y}_{\text{w.on}} \in \mathbb{R}^{8760\times 488}\,, \boldsymbol{y}_{\text{sol}} \in \mathbb{R}^{8760\times 489}\,,$$

for the available power output of wind offshore, $\boldsymbol{y}_{\text{w.off}}$, wind onshore, $\boldsymbol{y}_{\text{w.on}}$, and solar generation, $\boldsymbol{y}_{\text{sol}}$, for each hour of the year 2011. The authors select a portion of the data $\{1, \dots, N\}$, corresponding to the month of January, to be used to fit the models. For each time series, let $y_{(\cdot)}(t, j)$ denote the available power output in MW at time $t \in \{1, \dots, N\}$ for the j-th generator, where $(\cdot)$ is either wind or solar.

[1] The authors use the general linear programming kit (GLPK) solver. We note that a different solver may produce a different solution.

Wind model As a pre-processing step, the two time series $\boldsymbol{y}_{\text{w.off}}$ and $\boldsymbol{y}_{\text{w.on}}$ are merged by summing the onshore and offshore power generation at the buses $\mathcal{I}_{\text{w.off}} \subset \mathcal{I}_{\text{w.on}}$. This results in

$$y_w(t,j) = y_{\text{w.on}}(t,j) + \mathbb{1}_{j\in\mathcal{I}_{\text{w.off}}} y_{\text{w.off}}(t,j)\,,$$

where $\mathbb{1}_{j\in\mathcal{I}_{\text{w.off}}}$ is the indicator that wind farm j is located at a bus with an offshore wind farm. For each wind farm j, the power generation is modelled by an ARMA(1,24) such that

$$\begin{aligned} y_w(t,j) &= a_{1,j} y_w(t-1,j) + e_w(t,j) \\ &+ m_{1,j} e_w(t-1,j) + \cdots + m_{24,j} e_w(t-24,j)\,, \end{aligned} \tag{3.4}$$

where $a_{1,j}$ and $m_{k,j}$, $k = 1,\ldots,24$, are the estimated auto-regressive and moving average coefficients, respectively, and $e_w(1,j),\ldots,e_w(N,j)$ are the residuals.

Solar model The authors describe the model for solar generation as an ARMA(24,24) with all parameters fixed to 0, except for the ones corresponding to the seven hours before, and the one corresponding to twenty-four hours before. Each solar farm j is represented by

$$\begin{aligned} y_{\text{sol}}(t,j) &= a_{1,j} y_{\text{sol}}(t-1,j) + \cdots + a_{7,j} y_{\text{sol}}(t-7,j) \\ &+ a_{24,j} y_{\text{sol}}(t-24,j) + e_{\text{sol}}(t,j) \\ &+ m_{1,j} e_{\text{sol}}(t-1,j) + \cdots + m_{7,j} e_{\text{sol}}(t-7,j) \\ &+ m_{24,j} e_{\text{sol}}(t-24,j)\,, \end{aligned} \tag{3.5}$$

where $e_{\text{sol}}(1,j),\ldots,e_{\text{sol}}(N,j)$ are the residuals. The model captures the dependency on the hour of the day as well as the projection of the current day. We refer to the model as a *seasonal auto-regressive moving average* (SARMA) given by SARMA$(7,7)(1,1)_{24}$ where the seasonal parameters correspond to the 24 hour lag. This notation clearly identifies the model parameters that are estimated.

Estimating the covariance The variance of power generation is estimated as the empirical variance of the residuals $e_{(\cdot)}(t,j)$, $t = 1,\ldots,N$, for each generator of a given type, where $(\cdot)$ is either wind or solar. Denote $U_{(\cdot)} \in \mathbb{R}^{N\times n_s}$ by the residual matrix for a given generator type such that the ik-th entry is given by

$$(U_{(\cdot)})_{ik} = \mathbb{1}_{j\in k} e_{(\cdot)}(i,j)\,,$$

where $\mathbb{1}_{j\in k}$ is the indicator that generator j is located at bus k. The covariance matrix of a given generator type is given by

$$\boldsymbol{\Sigma}_{(\cdot)} = \frac{1}{N}\boldsymbol{u}^\top \boldsymbol{u}\,,$$

where $\boldsymbol{u} = U_{(\cdot)} - \frac{1}{N}\mathbf{1}\mathbf{1}^\top U_{(\cdot)}$, and $\mathbf{1}$ is an $N \times 1$ column vector of ones. The residual matrix for solar generation contains only residuals $e_{\text{sol}}(t,j)_{t\in\mathcal{D}}$, where $\mathcal{D}$ is the set of daylight hours, resulting in a residual matrix $U_{\text{sol}} \in \mathbb{R}^{N^*\times n_s}$ where $N^* < N$.

Assuming that the residuals for wind and solar are independent, the power injection covariance matrix for a given hour h is modelled by

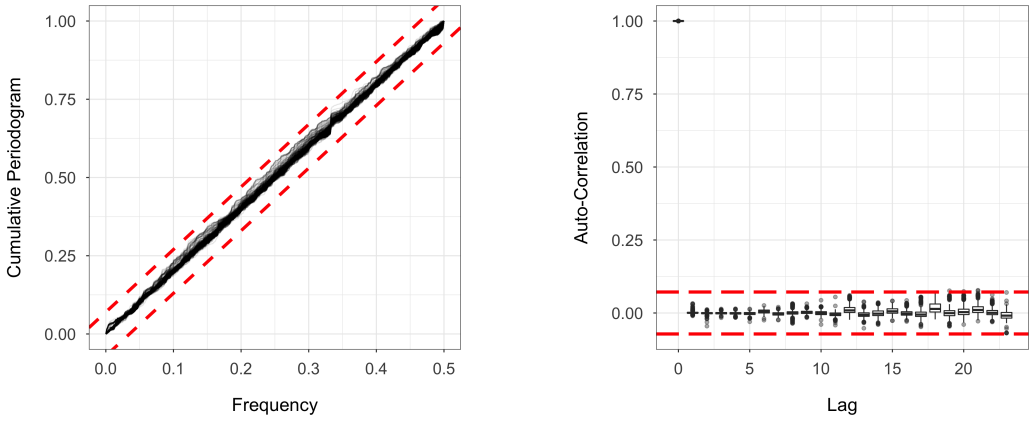
$$\Sigma_{p_s}(h) = \Sigma_w + \mathbb{1}_{h\in\mathcal{D}}\Sigma_{\text{sol}}\,,$$

where $\Sigma_w \in \mathbb{R}^{n_s\times n_s}$ is the estimated covariance matrix of wind generation, $\Sigma_{\text{sol}} \in \mathbb{R}^{n_s\times n_s}$ is the estimated covariance matrix of solar generation, and $\mathbb{1}_{h\in\mathcal{D}}$ is the indicator that the hour h corresponds to daylight.

3.2 Goodness-of-fit

In Section 3.1.6, we present the time series models of wind and solar generation, which the authors use to estimate the stochastic covariance matrix, Σ_{p_s}. These models may have a significant effect on the ranking of decay rates, δ_ℓ, and therefore we assess their goodness-of-fit using the cumulative periodogram and auto-correlation diagnostic plots from Section 2.4.5.

Wind model There are 488 time series of wind generation. The ARMA$(1, 24)$ is fitted to each time series. We present the goodness-of-fit diagnostic plots for the residual series of each model fit in Figure 3.2.



(a) Cumulative periodogram.

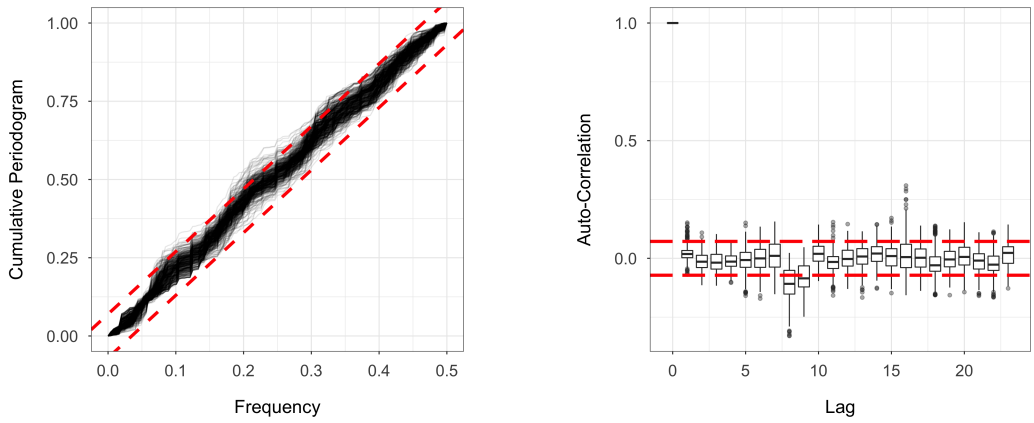
(b) Auto-correlation.

Figure 3.2: Goodness-of-fit diagnostic plots of the residuals from the ARMA$(1, 24)$ fitted to the 488 time series of wind generation. The dashed red lines indicate the 5% level of significance. (a) Each line corresponds to the cumulative periodogram of the residual series from one model fit. (b) Box-plots of the auto-correlation of the residual series for all the fitted models. From the diagnostics we conclude that the model is reasonable.

In Figure 3.2a, each cumulative periodogram is linear and remains within the 5% level of significance. Therefore, we conclude that no significant cycles are present. In Figure 3.2b, the box-plots are centred at zero and contained within the 5% level of significance. Therefore, we conclude that the residuals are not auto-correlated, and that the ARMA$(1, 24)$ is a reasonable model for wind generation.

Solar model There are 489 time series of solar generation. The SARMA$(7, 7)(1, 1)_{24}$ is fitted to each time series. We present the goodness-of-fit diagnostic plots for the residual series of each model fit in Figure 3.3.

The cumulative periodogram in Figure 3.3a, consists of 157 models that violate the 5% level of significance. Whilst the violation is relatively small, the cumulative periodogram displays three peaks that suggests the presence of significant cycles. Furthermore, the auto-correlation in Figure 3.3b demonstrates that the majority of models have a negligible auto-correlation. However, the mean auto-correlation at a lag of 8 hours falls below the 5% significance level. We conclude that the model does not explain all seasonal cycles, and extend the model in Section 3.3.



(a) Cumulative periodogram.

(b) Auto-correlation.

Figure 3.3: Goodness-of-fit diagnostic plots of the residuals from the SARMA$(7,7)(1,1)_{24}$ fitted to the 489 time series of solar generation. The dashed red lines indicate the 5% level of significance. (a) Each line corresponds to the cumulative periodogram of the residual series from one model fit. (b) Box-plots of the auto-correlation of the residual series for all fitted models. The cumulative periodogram is non-linear and indicates the presence of potentially seasonal cycles in the residuals.

3.3 Solar model extension

Solar generation is highly dependent on solar irradiance which is a measure of light energy from the Sun at the Earth's surface. The amount of light energy at the earth's surface depends on the time of day, the time of year as well as other atmospheric variables that produce cloud formations.

The time of day is a measure of Earth's rotation which usually results in daylight hours and night-time hours (periods of no sun). The time of year is a measure of Earth's elliptical orbit around the sun which creates seasons where Earth is closer or farther from the sun. These are deterministic cyclic behaviours that can be predicted, whereas predicting cloud formations is more difficult.

Boland [9] proposes the decomposition of solar power output, S_t, into a deterministic trend, D_t, and a stochastic component, ψ_t, such that

$$S_t = \begin{cases} D_t + \psi_t & t \in \mathcal{D}\,, \\ 0 & \text{otherwise}\,, \end{cases}$$

where $\mathcal{D}$ denotes daylight hours. Fourier series is used to model the deterministic

trend, and an ARMA(7, 7) is used to model the stochastic component. We remove the daily seasonal component of the SARMA model as this will be captured in the Fourier series if needed. We refer to this model as a *Fourier filtered auto-regressive moving average* (FFARMA) model.

3.3.1 Deterministic model

Spectral analysis is performed to identify significant cycles. In particular, we calculate the average power spectrum given by

$$\overline{v}(\omega) = \frac{1}{489} \sum_{j=1}^{489} v_j(\omega) \,,$$

where $v_j(\omega)$ is the power spectrum of $y_{sol}(t, j)$, $t = 1, \ldots, 8760$. Figure 3.4 plots the average power spectrum normalised such that values represent the contribution of variance. Then in Table 3.1, we rank the cycles based on their contribution to the variance of solar generation.

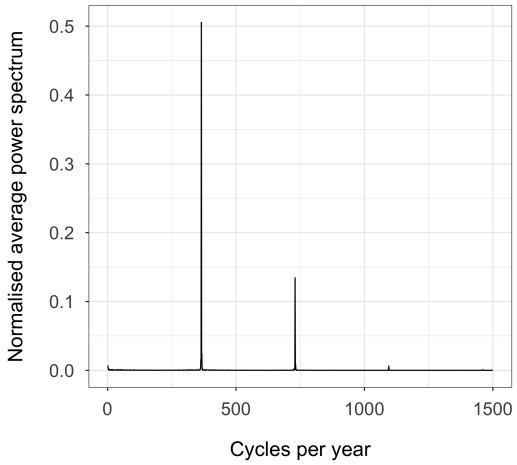


Figure 3.4: Normalised average power spectrum of solar generation for German solar farms. The x-axis is the number of cycles per year given by 8760ω, and the y-axis is the average power spectrum, $\overline{v}(\omega)$, normalised by total power. On average, the daily and twice-daily cycles explain over 60% of the variation in solar generation.

We choose to fit the Fourier series using the significant cycles corresponding to $1, 2, 365, 730$ and 1095 cycles per year, where 730 and 1095 cycles per year correspond to the twice and thrice daily cycles, respectively. Additionally, the ranking of the cycles, in Table 3.1, demonstrates that up to two side-bands have significant power. Side-bands modulate the width of the cycle to account for the varying lengths of

day between summer and winter. The deterministic Fourier series model is given by

$$\begin{aligned} D_t &= \alpha_0 + \alpha_1 \cos\frac{2\pi t}{8760} + \beta_1 \sin\frac{2\pi t}{8760} + \\ &\quad \alpha_2 \cos\frac{4\pi t}{8760} + \beta_1 \sin\frac{4\pi t}{8760} + \\ &\quad \sum_{i=3}^{17}\sum_{n=1}^{3}\sum_{m=-2}^{2}\left[\alpha_i \cos\frac{2\pi(365n+m)t}{8760} + \beta_i \sin\frac{2\pi(365n+m)t}{8760}\right], \end{aligned}$$

where α_i and β_i are the Fourier series coefficients. On average, the deterministic model explains over 75% of the total variation. We provide an example of the fit of the deterministic model to the actual solar generation in Figure 3.5.

Cycles per year	Normalised $\bar{G}_j(\omega)$
365	0.506
730	0.134
366	0.025
367	0.023
364	0.018
363	0.015
732	0.009
1	0.007
1095	0.006
2	0.005

Table 3.1: The top 10 cycles per year with the highest contribution to the variance. The side-bands of the daily cylce, *e.g.* 364 and 366, are also significant.

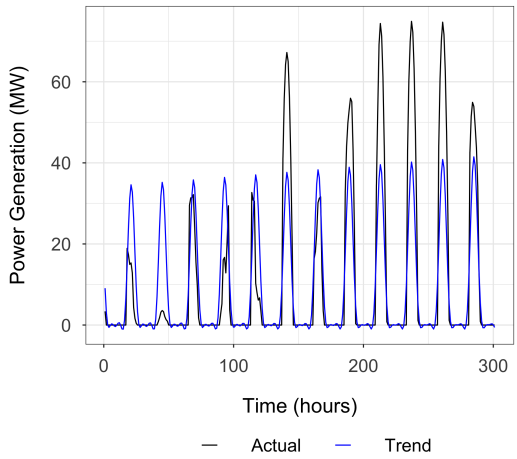
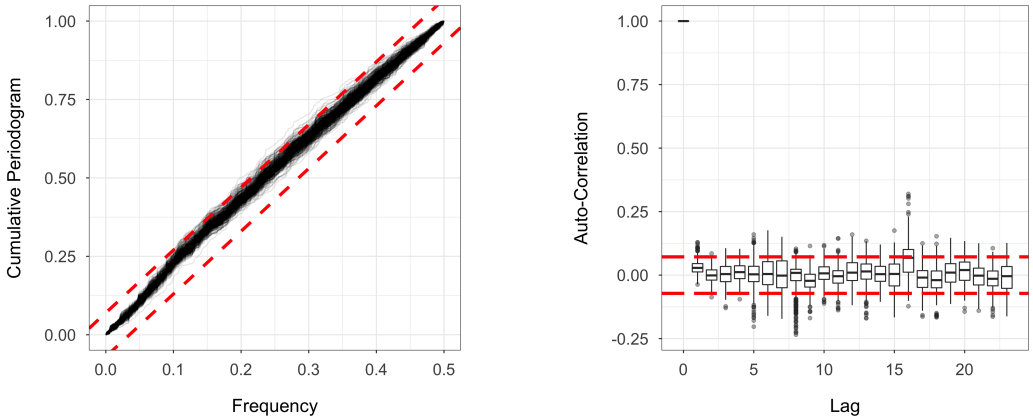


Figure 3.5: An example time series of solar generation (black) overlaid with the deterministic trend (blue), D_t. The deterministic trend predicts the average daily profile but does not explain the day-to-day or intra-day variations.

3.3.2 Goodness-of-fit

In the FFARMA model, prior to fitting the stochastic component, X_t, the deterministic model, D_t, on average explains over 75% of the variance in solar generation. This means that the stochastic ARMA$(7, 7)$ model is fitted to a less variable time series. Therefore, we expect the diagnostic plots of the FFARMA model in Figure 3.6, will have an improved fit.



(a) Cumulative periodogram.

(b) Auto-correlation.

Figure 3.6: Goodness-of-fit diagnostic plots of the residuals from the FFARMA model fitted to the 489 time series of solar generation. The dashed red lines indicate the 5% level of significance. (a) Each line corresponds to the cumulative periodogram of the residual series from one model fit. (b) Box-plots of the auto-correlation of the residual series for each fitted model. The diagnostics demonstrate an improved model fit.

The cumulative periodogram in Figure 3.6a, is linear and only contains 47 models that violate the 5% level of significance. Whilst this is not perfect, the violations appear negligible and therefore we conclude that no seasonality is present in the residuals. The auto-correlation in Figure 3.6b, consists of zero mean box-plots at all lags except at a lag of 16 hours and is a minor improvement upon the previous model. Therefore, we conclude that the FFARMA model has an improved goodness-of-fit over the SARMA model used by Nesti *et al.* [38].

3.3.3 Results

Nesti *et al.* [38] report the ranking of the most vulnerable lines in the German power network, relative to the hour 11 am of the day 01/01/2011. We reproduce the results in Table 3.2a[2]. Then, we perform the same method, using the FFARMA model to estimate the solar covariance matrix, Σ_{sol}. The new ranking is presented in Table 3.2b.

ℓ	$\mathbb{P}(\lvert f_\ell \rvert \geq 1)$	I_ℓ
361	1.744×10^{-2}	2.226
803	8.228×10^{-4}	4.954
19	6.784×10^{-4}	5.132
27	6.033×10^{-4}	5.24
670	3.527×10^{-4}	5.737
809	7.575×10^{-5}	7.177
586	5.574×10^{-5}	7.466
587	5.454×10^{-5}	7.486
810	2.496×10^{-5}	8.225
389	2.449×10^{-5}	8.244
390	2.420×10^{-5}	8.255
712	6.440×10^{-6}	9.514
682	5.337×10^{-6}	9.693
683	5.318×10^{-6}	9.697
714	3.876×10^{-6}	9.999
715	1.052×10^{-6}	11.249
554	4.267×10^{-7}	12.117
488	4.209×10^{-7}	12.13
707	1.199×10^{-7}	13.341
818	1.199×10^{-7}	13.341

(a) Reproduced results from [38].

ℓ	$\mathbb{P}(\lvert f_\ell \rvert \geq 1)$	I_ℓ
361	1.693×10^{-2}	2.251
27	8.454×10^{-5}	7.073
670	7.656×10^{-5}	7.167
19	3.340×10^{-5}	7.95
389	2.048×10^{-5}	8.413
390	2.024×10^{-5}	8.424
803	1.083×10^{-5}	9.019
809	5.982×10^{-6}	9.585
586	4.909×10^{-6}	9.773
587	4.773×10^{-6}	9.8
516	3.026×10^{-6}	10.236
810	2.108×10^{-6}	10.582
712	4.143×10^{-7}	12.145
554	2.796×10^{-7}	12.524
488	2.757×10^{-7}	12.538
714	2.409×10^{-7}	12.668
682	1.022×10^{-7}	13.496
683	1.017×10^{-7}	13.5
707	2.812×10^{-8}	14.744
818	2.812×10^{-8}	14.744

(b) Σ_{sol} estimated from the FFARMA model.

Table 3.2: Line indexes, exact failure probabilities, and decay rates for the 20 top most vulnerable lines, 11 am. (a) Solar covariance matrix, Σ_{sol}, is estimated from the SARMA model from [38]. (b) Solar covariance matrix, Σ_{sol}, is estimated using our FFARMA model. The ranking of decay rates changes when we modify the solar model.

It is evident in Table 3.2, that the ranking is sensitive to the models used for es-

[2]The reproduced ranking differs to the ranking presented in [38]. We find that lines 389 and 390 are not operating at capacity at 11 am and thus, are ranked lower. This is also the case in the code provided to us by the authors.

timation. Whilst line 361 remains the most vulnerable, almost every other line is in a new position. Still, it is worth noting that most (if not all) of the original most vulnerable lines make it to the new list. Additionally, the reported decay rates corresponding to the FFARMA model are larger than the original decay rates in Table 3.2a and therefore exact failure probabilities have decreased.

3.4 Discussion

In this chapter, we discussed in-depth, the emergent cascade model of Nesti *et al.* [38] and its application to the German power network, including the models of wind and solar generation. Our goodness-of-fit analysis of the wind model fitted to the 488 time series of wind generation, demonstrated no significant cycles or auto-correlation. Therefore, we concluded that the wind model was reasonable. On the other hand, we found that 157 models of the 489 time series of solar generation violated the 5% critical level of the cumulative periodogram. Additionally, the shape of all cumulative periodograms were non-linear and suggested the presence of significant cycles, and motivated our extension.

In Section 3.3, our extension to the solar model follows the ideas of Boland [9], and considers the deterministic and stochastic components of solar generation separately. This allowed us to fit a deterministic Fourier series to the significant cycles identified by the power spectrum. The identified Fourier series explained, on average, over 75% of the variation in solar generation. The goodness-of-fit of our extended FFARMA model contained only 47 models that violated the 5% critical level of the cumulative periodogram. Whilst this is still not perfect, the cumulative periodograms were significantly more linear in shape. Therefore, we concluded our FFARMA model had an improved fit.

Finally, we compared the line indexes, exact failure probabilities and decay rates for the 20 top most vulnerable lines at 11 am. We found that the most vulnerable line, $\ell = 361$, reported by Nesti *et al.* [38], remained the most vulnerable after extending the solar model. Since the authors assume the most vulnerable line is the first to failure, our extension does not affect their results. However, using the FFARMA model increased the exact failure probabilities across all lines, and would impact a simulation-based study.

The authors debate the main assumption of the emergent cascade model which is that fluctuations in renewable energy follows a multivariate Gaussian distribution.

We had considered that technical features of wind turbines such as brakes and pitch control, motivate a *truncated multivariate Gaussian distribution* that is bounded above such that

$$\boldsymbol{p} \sim TN_n(\boldsymbol{\mu}, \Sigma_p) \quad \text{s.t. } \boldsymbol{p} \leq \boldsymbol{b}$$

where $\boldsymbol{b}$ is the column vector of upper limits. We recognised that f_ℓ is a linear combination of the power injections, $\boldsymbol{p}$, given by

$$f_\ell = \boldsymbol{a}^\top \boldsymbol{p}$$

where $\boldsymbol{a}$ is a column vector of the ℓth row of the matrix of PTDFs, H. Therefore, the marginal distribution of the power flow along line ℓ is characterised by the moment generating function given by

$$\begin{aligned} M_{f_\ell}(t) &= \mathbb{E}[\exp(\boldsymbol{a}^\top \boldsymbol{p} t)] \\ &= \frac{\int_{-\infty}^{\boldsymbol{b}-\boldsymbol{\mu}} \exp(-\frac{(\boldsymbol{a}^\top \boldsymbol{w} - \sigma^2 t)^2}{2\sigma^2})\ d\boldsymbol{w}}{\int_{-\infty}^{\boldsymbol{b}-\boldsymbol{\mu}} \exp(-\frac{\boldsymbol{a}^\top \boldsymbol{w}}{2\sigma^2})\ d\boldsymbol{w}} \end{aligned}$$

where the transformation $\boldsymbol{w} = \boldsymbol{p} - \boldsymbol{\mu}$ is applied for simplification [27]. This is not a known distribution and must be solved numerically. This is a complicated problem and is left as future work.

Another point to note, is that the stochastic covariance matrix, Σ_{p_s}, is estimated from hourly data, whereas power networks dispatch generation on shorter time intervals such as 5-15 minutes. The forecast error of semi-scheduled generation decreases for shorter time frames and therefore we expect the exact failure probabilities to decrease further. It would be interesting to see the impact of a smaller data resolution on the magnitude of the exact failure probabilities.

Most importantly, the emergent cascade model needs to be validated. We suggest to identify cascades from real unplanned outage data and determine if any exhibit average line propagations similar to emergent cascades. Depending on the case study, unplanned outage data may be very difficult to obtain. We had intended to perform a case study of the emergent cascade model on the South Australian power network which included simulation and validation. However, conversations with AEMO and ElectraNet suggested that obtaining unplanned outage data is highly unlikely. Given our strong intentions to perform a case study on the SA power network, we left this as future work. This concludes our discussion of emergent cascades, as we now divert our focus to network congestion.

Chapter 4

South Australian power network

The first wind farm built in SA was the 34 MW Starfish Hill wind farm commissioned in September 2003. At the end of 2019, the total installed capacity of wind generation in SA exceeded 1,800 MW. This combined with SA's relatively small average demand has meant that within the NEM, SA has the highest percentage of wind generation, as seen in Figure 4.1.

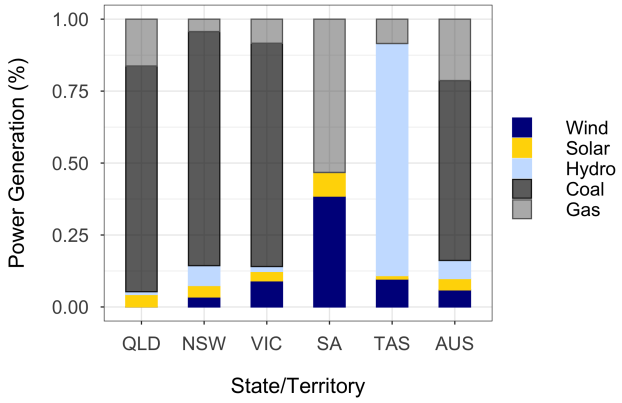


Figure 4.1: Percentage of power generation for each fuel type in each state for the 2017-18 financial year [42]. AUS refers to the overall fuel mix of Australia. In SA, wind power contributed almost 40% of all electricity generation.

In Chapter 6, we analyse the network congestion that would result from increasing installed wind capacity in the South Australian power network. The analysis performs a simulation study of the network power flows, the calculations of which require a model of the SA power network. Ideally, the model would be an exact representation of the real network. However, the information needed for an exact representation is not available. Therefore, we develop a simplified network model

from publicly available data.

The network model is constructed of two parts: the physical model and its dynamic inputs. The physical model represents the real physical network components, whereas the dynamic inputs correspond to the scheduling of generation to meet a varying demand. The physical model and its dynamic inputs determine the network power flows.

In this chapter, we outline a simplified open-access AC model of the SA power network for the year 2015. We first estimate the physical model for the real network. Then, in Section 4.2, we determine realistic inputs for 2015. In Section 4.3, we perform model diagnostics to ensure that the AC model reflects realistic operation. Then in Section 4.4, we describe a DC approximation to the AC model and assess its accuracy. We conclude this chapter with a brief visualisation of the power transmission distribution factors.

4.1 Physical model

AC power flow calculations require more information than the DC model. A lot of this information is not published. Therefore, we assume realistic values for the parameters when there is no published data. In this section, we discuss how each parameter for each network element is estimated and include the source where applicable. Currently, our model does not include synchronous condensers. Table 4.1 provides an overview of the parameters and whether they are publicly available.

4.1.1 Topology

The network topology is a list of the buses, branches, generators, loads and interconnectors, and consists of their connections. The topology is mapped from AEMO's 2019 high voltage main systems diagram [53], which is publicly available. However, the publicly available dynamic inputs (as discussed in Section 4.2) do not contain industrial loads. Therefore, we simplify the topology by combining the buses presented in Appendix A, and a plot of the resulting network is shown in Figure 4.2. Additionally, the edge-vertex incidence matrix is provided in Appendix A.2.

Element	Parameter (units)	Variable	Published
Bus	Voltage level (kV)	V	✓
Branch	Capacity (MW)	c	✓
Branch	Resistance (Ω)	R	✗
Branch	Inductive reactance (Ω)	X	✗
Branch	Shunt susceptance (S)	B_c	✗
Branch	Length (km)†	-	✓
Generator	Active power capacity (MW)‡	P_{max}	✓
Generator	Maximum reactive power limit	Q_{max}	✗
Generator	Minimum reactive power limit	Q_{min}	✗
Generator	Power factor	φ_G	✗
FACTS	Static Var Compensator (SVC) location	-	✓
FACTS	Maximum SVC reactive power limit	Q_{max}	✗
FACTS	Minimum SVC reactive power limit	Q_{min}	✗
FACTS	Shunt capacitance power factor	φ_C	✗

Fast AC Transmission System (FACTS)

† indicates a parameter that is not directly used for load-flow analysis.

‡ we model inter-connectors as generators.

Table 4.1: Overview of the network parameters.

4.1.2 Buses

The network buses are identified from the topology and are characterised by their voltage level. The SA transmission network is defined on two voltage levels: 132 kV and 275 kV. AEMO's high voltage main systems diagram [53] identifies the voltage level of each bus. A bus that houses a transformer is split into two buses defined on each voltage level. For example, the bus Para is split into Para 132 and Para 275, and there exists a transformer that connects Para 132 to Para 275.

There are three types of network buses:

1. **Generator bus** - has at least one generator attached.
2. **Load bus** - any bus that has no power generation capability.
3. **Slack bus** - designed to make up for transmission losses. In the model, we al-

locate the slack to the South East 275 kV bus, as this corresponds to significant import/export between SA and Victoria.

The topology consists of 25 generator buses, 16 load buses and one designated slack bus.

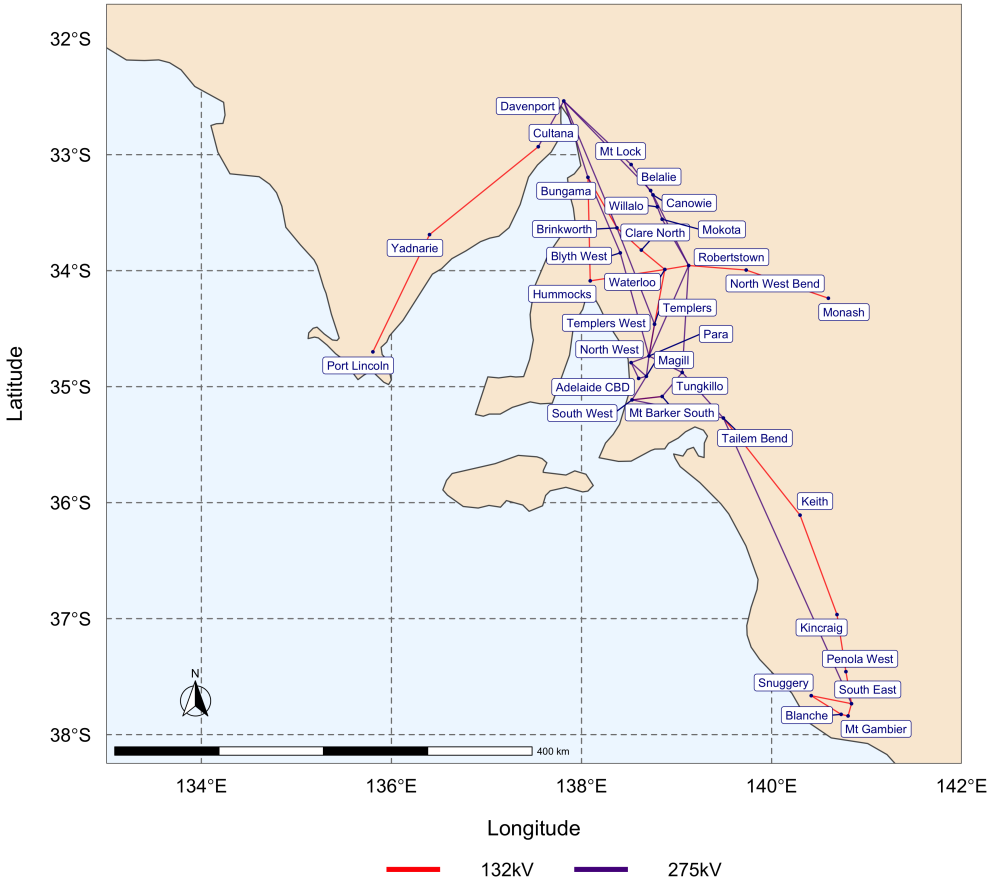


Figure 4.2: Map of the simplified South Australian power network topology with 42 buses, 49 branches and 7 transformers. The edge colour indicates the transmission line voltage on two levels: 132 kV and 275 kV. The network inter-connects to Victoria via transmission lines connected to the Monash and South East buses.

4.1.3 Branches

The branches consist of 49 transmission lines and 7 transformers. The branch parameters are the most important to accurately estimate, as they determine the maximal power flows and transmission losses. The branch parameters consist of the capacity

c, resistance R, inductive reactance X and shunt susceptance B_C.

Branch Capacity

The branch capacity is the maximal power transfer across a branch. Therefore, it is critical for congestion analysis as it determines the point at which the line becomes congested. The branch capacities are published in AEMO's transmission equipment ratings dataset [46]. These are apparent power ratings measured in MVA, which include both active and reactive power. Some transmission lines are dynamically rated such that the different seasons, spring/autumn, summer and winter, each has a corresponding branch capacity. We use the dynamically rated capacities to determine congestion.

Transmission Parameters

There is no public data available for the resistance R, inductive reactance X and shunt susceptance B_C of the transmission branches. Therefore, we estimate the parameters of the transmission lines from their length and the parameters of the transformers from their capacity.

The Australian Government's data repository includes the *Electricity Transmission Lines* dataset produced by Geoscience Australia in 2015 [1]. This dataset provides the transmission line lengths in decimal degrees (dd). We convert this into kilometres (km) by using 1 dd = 111.32 km. Lines that are not included in the dataset, are approximated by straight line distances between the *to* and *from* buses. The transmission parameters of the lines are estimated using typical per kilometre values, displayed in Table 4.2.

Parameter	Units	132 kV	275 kV
R	Ω / ph / km	0.112	0.040
X	Ω / ph / km	0.39	0.319
B_C	S / ph / km	2.564×10^{-6}	3.651×10^{-6}

Electrical resistance is measured in ohms, $\Omega = \frac{V}{A}$, where V is voltage and A is amperes.

Susceptance is measured in Siemens, $S = \frac{1}{\Omega}$.

ph refers to a per phase representation, which is converted to per-unit for use in load-flow analysis.

Table 4.2: Typical overhead transmission line parameters (rated frequency 50 Hz).

The transformers are defined on their reactance such that the resistance $R = 0$ and the shunt capacitance $B_C = 0$. The transformer reactance X is given by

$$X = 0.12 \frac{P_{base}}{c} ,$$

where c is the branch capacity and P_{base} is the power base for per-unit calculations.

4.1.4 Generators

There are two distinct types of generation data: network parameters and time series of generation. We discuss the network parameters here, and the time series data later in Section 4.2.

Reactive power limits

All generators have active and reactive power capacity limits (or rated power). The active power capacity limit is obtained from AEMO's generation information [51], and is used to calculate the reactive power limits. The maximum reactive power limit, Q_{max}, of a generator is proportional to the generation capacity, P_{max}, such that

$$Q_{max} = P_{max} \tan(\cos^{-1}(\varphi_G)) ,$$

where φ_G is the rated power factor. We choose a conservative rated power factor $\varphi_G \approx 0.95$. The minimum reactive power limit, Q_{min}, is normally a larger magnitude than the maximum, as the generators are required to absorb reactive power in the network. However, when congestion is present the system is under stress and only small amounts of reactive power are absorbed, and thus we approximate

$$Q_{min} = -Q_{max} .$$

Connection

When calculating DC power flows, it is reasonable to have the generators located at the high voltage buses because the DC approximation assumes the voltages are 1 pu. However, when calculating AC power flows, the voltages may vary based on the network loading. Therefore, we connect the generators to the high voltage buses via transformers. The transformers have the ability to boost the voltage level to above

1 pu at the high voltage bus. This is common practice in power systems engineering as operating voltage levels above 1 pu decreases transmission losses.

Conventional generators are connected by one transformer with reactance

$$X_{G_1} = 0.12 \frac{P_{base}}{P_{max}},$$

where P_{max} is the generators rated power and P_{base} is the power base for per-unit calculations. Wind farms are connected by two transformers to allow the voltage of the intermediary bus to be controlled to 1 pu. We assume the reactance X_{W_1} and X_{W_2} are given by

$$\begin{aligned} X_{W_1} &= 0.05 \frac{P_{base}}{P_{max}}, \\ X_{W_2} &= 0.1 \frac{P_{base}}{P_{max}}. \end{aligned}$$

The capacity of these transformers are given by

$$c_{trans} = \frac{P_{max}}{\varphi_{trans}},$$

where $\varphi_{trans} = 1.01$ is the transformer power factor. Most importantly, the transformers attached to the high voltage buses are equipped with a tap ratio of 1.05. This means that a voltage level of 1 pu at the *from* bus is increased to 1.05 pu at the *to* bus.

4.1.5 Inter-connectors

South Australia is connected to Victoria via two inter-connectors: Heywood and Murraylink. At any given time the inter-connectors are either importing power from Victoria or exporting excess power to Victoria. We represent the inter-connectors as a generator, with effectively infinite capacity, attached to the corresponding high voltage bus in SA. Therefore, power imports across the inter-connector are represented as generation and power exports are represented as negative generation. The benefit to modelling the inter-connectors as generators, is that negative generation, which is effectively demand, still produces reactive power to support voltages.

Since the inter-connectors are being modelled as generators, the corresponding voltage set point is chosen to be 1.03 pu. This means that the SA buses South East and Monash connected to Heywood and Murraylink, respectively, will have a fixed voltage of 1.03 pu. Whilst a fixed voltage is unrealistic, this assumption means it is guaranteed to be greater than 1 pu, in order to support the power transfer across the inter-connectors.

4.1.6 Flexible AC transmission system devices

Flexible AC transmission system (FACTS) devices are static electrical components of the network designed to provide voltage control and enhance the power transfer capability. FACTS devices can be connected in series with the power network (series compensation) and in shunt with the power network (shunt compensation) [37]. Whilst power networks include FACTS devices, there is minimal reporting on their locations and specific parameters.

We incorporate shunt capacitive compensation into the model to improve the power factor at load buses. In the real network, capacitors are located at the connections to the distribution network. When demand is high, the capacitors are turned on in order to increase the power factor. We attempt to emulate this by turning on shunt capacitance when the power factor drops below 0.9. This is designed to achieve a more realistic power factor of 0.98.

In addition to capacitors, static VAR compensators (SVCs) are typically connected to the power network near large loads. These electrical components can both produce and absorb large amounts of reactive power. The SA network has two SVCs; one at Para and one at South East [44]. We model the SVC at the Para 275 kV bus as a generator with 300 MW of reactive power and no active power. Since we are modelling the South East inter-connector as a generator, the additional reactive power that the generator provides emulates the operation of the SVC as well.

4.2 Dynamic inputs

In the operation of the power network, generation is scheduled every five minutes in order to meet demand. This is a complex optimisation designed to minimise cost whilst ensuring that the network constraints are satisfied. The generation values for each generator and the demand at each load are the inputs to the physical model, which together determine the power flows.

We describe below the time series generation and load data for the AC model of the year 2015. We choose to use the time series data corresponding to 2015 in order to avoid wind curtailment that has occurred due to limitations brought in after the state-wide blackout that occurred in September 2016 [43].

4.2.1 Power generation time series

The power generation time series of each generator in SA is sourced from AEMO's Market Management System (MMS) data model [50]. The MMS is the large data repository where all publicly available electricity market data is archived. We access the Supervisory Control and Data Acquisition (SCADA) dispatch data and interconnector flow data to compile our generation time series. Recall, that the interconnectors are modelled as generators which is why they are included here.

Dispatch data

The SCADA dispatch data is the measurement of a generator's output at the transmission connection point and is available for every 5-minute interval. A generator that is not dispatched will either have no power generation and be turned off, or consume power to run essential equipment. We will see this later in Chapter 5 when investigating wind generation.

The dataset contains missing data, which is distributed among the generators as indicated by Table 4.3. Since the amount of missing data is negligible, we perform linear imputation. This is a linear interpolation between the two observations at the boundary of the missing region. The linear assumption fails as the interval length increases. Fortunately, the median length of the missing regions is one data point, and the mean length is three consecutive missing data points. Therefore, the imputation method is reasonable.

ID	Missing data
CATHROCK	0.09%
CNUNDAWF	0.59%
LKBONNY1	0.01%
MTMILLAR	0.05%
STARHLWF	0.51%
WPWF	0.04%

Table 4.3: Percentage of missing data within the time series of generation for each generator. Generators that are not listed have a complete time series.

Inter-connector flow data

The inter-connector power flow data is a measurement of the power transfer across the corresponding inter-connecting transmission line. The MMS publishes this data at 30-minute intervals [50], which is a lower resolution than the 5-minute dispatch. We perform linear interpolation to up-sample the inter-connector data to a 5-minute resolution to be compatible with the generation data. Whilst the change in inter-connector flows is relatively constrained, it may not necessarily be linear over a 30-minute period. We account for such inaccuracies in Section 4.2.3.

4.2.2 Demand time series

The distribution network acts as the loads of the transmission network. SA Power Networks (SAPN) manage the distribution network and publish *Zone Substation data* [39]. We use the Zone Substation data to create the demand time series for the loads of the transmission network.

The dataset contains the active and reactive demand recorded every 30-minutes at each distribution connection point. The distribution connection points have a corresponding associated transmission connection point. The total active and reactive demand at the transmission connection point is used as the active and reactive load time series. However, the distribution system in Adelaide may be connected across multiple transmission connection points. In particular, the dataset contains non-existent transmission connections points such as Metro North, Metro East, Metro West and Metro South. We allocate the demand to our simplified network as follows: Metro North to Para 275, Metro West to North West, Metro South to South West and the allocation for Metro East is displayed in Table 4.4.

The dataset does not include power that is supplied directly from the transmission network to industry, and some substations are excluded due to confidentiality [40]. Therefore, this dataset underestimates the total demand, which we address in Section 4.2.3.

Similarly to the inter-connector flows, we perform linear interpolation to up-sample the demand data to a 5-minute resolution to be compatible with the generation data. In this case, demand does not significantly change in a 30-minute period and thus linear interpolation is reasonable.

Magill	North West	Para 275
Burnside	Kilburn	Clearview
Campbelltown	North Adelaide	Golden Grove
Kent Town	Northfield	Harrow
Linden Park	Prospect	Hillcrest
Norwood		Holden Hill
Woodforde		Ingle Farm
		Hope Valley
		Tea Tree

Table 4.4: Allocation of Metro East demand from the distribution substation to the associated transmission connection point.

4.2.3 Balancing

The dynamic inputs are formed from a combination of sources, namely generation obtained from AEMO and load obtained from SAPN. This results in a significant discrepancy between total generation and total load. Such imbalances cause the Newton-Raphson algorithm for the AC power flows to be non-convergent, potentially because there is no feasible solution. Therefore, the inputs require generation and load balancing to obtain realistic inputs that account for transmission losses.

The dynamic inputs of generation and load form a set of power injections, $\boldsymbol{p}$, where the power injection at bus k is given by

$$p_k = g_k - d_k \,,$$

where g_k and d_k are the total generation and total demand at bus $k = 1, \ldots, n$, respectively. We form the raw power injections using the power generation time series from Section 4.2.1, the demand time series from Section 4.2.2, and assume a constant 200 MW demand at the Davenport bus for power supplied to the Olympic Dam and Prominent Hill mining operations. The algorithm in Figure 4.3 is used to balance the raw power injections in order to obtain the power injections used for the AC model.

Transmission Losses Unlike DC power which assumes a lossless network, AC power experiences power losses. Thus, in most cases total generation is greater than total demand to account for losses across the network. Therefore, we account for transmission losses, $\chi = 4\%$, when balancing the inputs. Variations in transmission losses are accounted for by the slack generator.

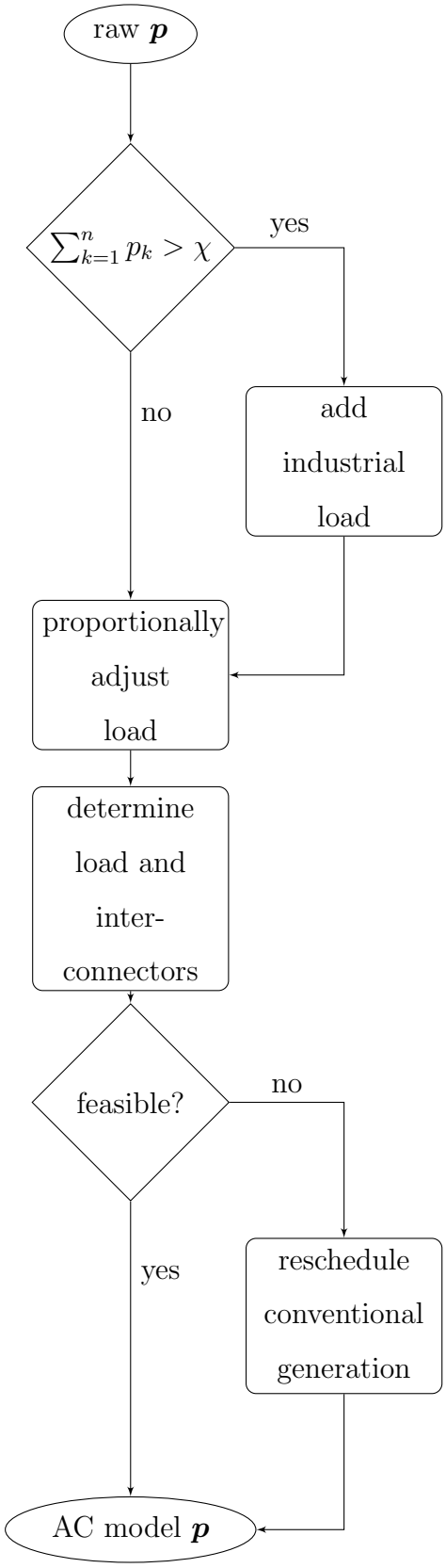


Figure 4.3: Flowchart of the algorithm to calculate the power injections for the AC model. The input is the raw power injections as described above. The output is the modified power injections used as the inputs for the AC model. This algorithm does not change the generation from wind farms.

Add industrial load As we described previously, the load data from SAPN did not include power supplied to industry direct from the transmission network. Therefore, we allocate load to major commercial and industrial sites as per Table 4.5. The total amount of industrial load is capped at 250 MW to respect a reasonable maximum industrial demand.

Bus	Contribution
Adelaide CBD	10%
Bungama	10%
Mt Barker	10%
North West	40%
South West	10%
Cultana 132	10%

Table 4.5: Contribution of industrial load.

Proportionally adjust load The central load buses are proportionally changed in order to achieve a realistic balance that accounts for transmission losses. The central buses include Adelaide CBD, Magill, Mt Barker South, North West, Para and South West. We restrict to adjusting the central loads because large imbalances can lead to unrealistically high rural demand.

Determine load and inter-connectors At this point, we have either increased or decreased load in order to obtain a realistic balance with the current generation profile. However, the previous processes are provisional and do not necessarily lead to power flows that respect network constraints. Therefore, we optimise the load and inter-connectors to find a solution that satisfies the constraints.

Suppose the power flows that result from the power injections at this point in the algorithm, $\boldsymbol{p}$, do not respect their transmission limits. Then, we apply a quadratic optimisation in order to determine a new set of power injections, $\boldsymbol{p}'$, such that the new power flows respect their transmission limits. We separate the power injections, $\boldsymbol{p}$, into two components: a variable component, $\hat{\boldsymbol{p}}$, and a fixed component, $\bar{\boldsymbol{p}}$. The variable component, $\hat{\boldsymbol{p}}$, consists of the load buses and the buses that correspond to the two inter-connectors. Thus, the fixed component, $\bar{\boldsymbol{p}}$, consists of generator buses that correspond to conventional and renewable generation. We optimise over

the variable component which reduces the problem dimension and the number of constraints.

The optimisation objective function minimises the change in the inputs given by

$$\min_{\hat{p}_k} \sum_k (\hat{p}'_k - \hat{p}_k)^2 \,.$$

Intuitively, the value of the objective function is smaller for solutions that have lots of small changes and larger for solutions that might only have one large change.

The AC power flows are non-linear and do not have a closed-form solution. Therefore, we constrain the power flows assuming a DC power flow model. Whilst DC power flows may not be as accurate, the optimisation can still inform a new generation solution to be used in AC power flow calculations. Denote by $\hat{H}$ and $\bar{H}$ the matrix of PTDFs restricted to the variable and fixed buses, respectively. The transmission constraint $|\boldsymbol{f}'| \leq \mathbf{1}$ is given by

$$|H\boldsymbol{p}'| \leq 1 \,,$$

which implies

$$-1 \leq \hat{H}\hat{\boldsymbol{p}}' + \bar{H}\bar{\boldsymbol{p}}' \leq 1 \,,$$

and therefore

$$\begin{aligned} \hat{H}\hat{\boldsymbol{p}}' &\leq 1 - \bar{H}\bar{\boldsymbol{p}} \,, \\ -\hat{H}\hat{\boldsymbol{p}}' &\leq 1 + \bar{H}\bar{\boldsymbol{p}} \,, \end{aligned}$$

where $\bar{\boldsymbol{p}}' = \bar{\boldsymbol{p}}$ because the fixed component remains constant. Additionally, the new power injections, $\boldsymbol{p}'$, must satisfy $\mathbf{1}^\top \boldsymbol{p}' = 0$, because the DC approximation assumes a lossless network. The resulting quadratic optimisation is given by

$$\begin{aligned} \min_{\hat{p}_k} \quad & \sum_k (\hat{p}'_k - \hat{p}_k)^2 \\ \text{s.t.} \quad & \hat{H}\hat{\boldsymbol{p}}' \leq 1 - \bar{H}\bar{\boldsymbol{p}} \,, \\ & -\hat{H}\hat{\boldsymbol{p}}' \leq 1 + \bar{H}\bar{\boldsymbol{p}} \,, \\ & \mathbf{1}^\top \hat{\boldsymbol{p}}' = -\mathbf{1}^\top \bar{\boldsymbol{p}} \,, \\ & \hat{\boldsymbol{p}}' \leq \hat{\boldsymbol{g}}_{avail} \,, \\ & \hat{\boldsymbol{p}}' \geq -\hat{\boldsymbol{d}}_{max} \,. \end{aligned} \tag{4.1}$$

where $\hat{\boldsymbol{d}}_{max}$ is the column vector of maximum demand and $\hat{\boldsymbol{g}}_{avail}$ is the column vector of available generation at each bus corresponding to the variable component. The optimisation presented here is similar to the short-term re-dispatch models in Section 6.1.

Reschedule conventional generation We extend the variable component of the power injections in the optimisation problem (4.1) to include generator buses that correspond to conventional generation. Renewable generators are treated as part of the fixed component, $\bar{\boldsymbol{p}}$, so that wind generation is preserved for simulation in Chapter 5.

4.3 Model diagnostics

Sections 4.1 and 4.2 form our AC model of the SA power network. Ideally, this should be representative of the real network and its operation such that:

- voltage levels must not operate outside of $(0.9, 1.1)$ per-unit and are operated within a soft limit of $(0.95, 1.05)$ per-unit, and
- branch flows should not exceed capacity.

The diagnostic plots of the simplified open-access AC model are displayed in Figures 4.4 and 4.5.

The bus voltages in Figure 4.4 demonstrate that voltages do not operate outside $(0.9, 1.1)$ per-unit and are typically restricted to $(0.95, 1.05)$ per-unit. It is quite reasonable that a few buses operate above 1.05 pu, in order to support the voltage at connected buses. On the other hand, voltage drops below 0.95 pu would be avoided wherever possible. It can be seen that the number of instances where voltage is below 0.95 pu in the AC model is negligible.

The branch power flows, from Figure 4.5, demonstrate only a few instances of branches operating outside their capacity. This includes the South East transformer and the Robertstown-Northwest Bend transmission line. In both cases, the violation is negligible. Therefore, we conclude that our AC model respects realistic operation.

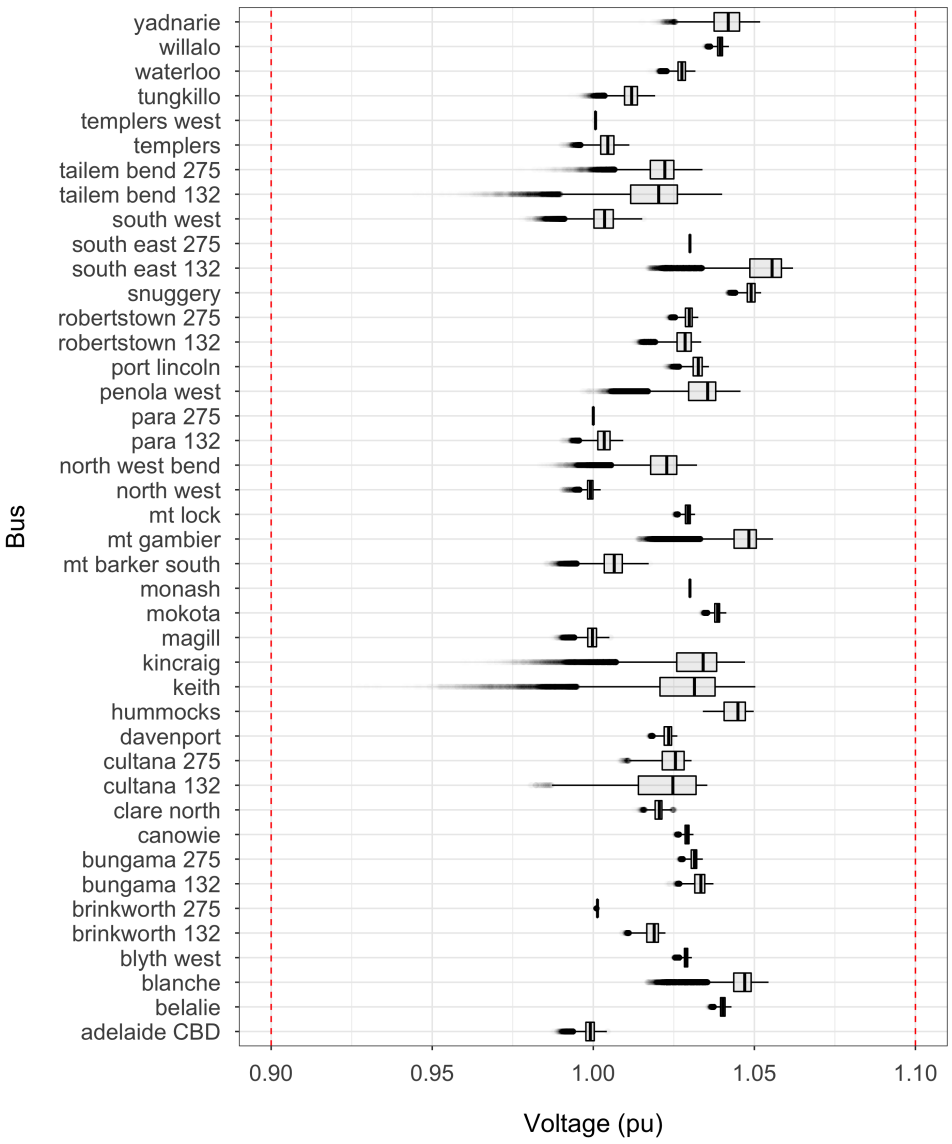


Figure 4.4: Diagnostic box-plots of the bus voltage levels (pu) for the AC model of the SA power network. The voltage levels of all the buses are contained within the limits $(0.9, 1.1)$ per-unit.

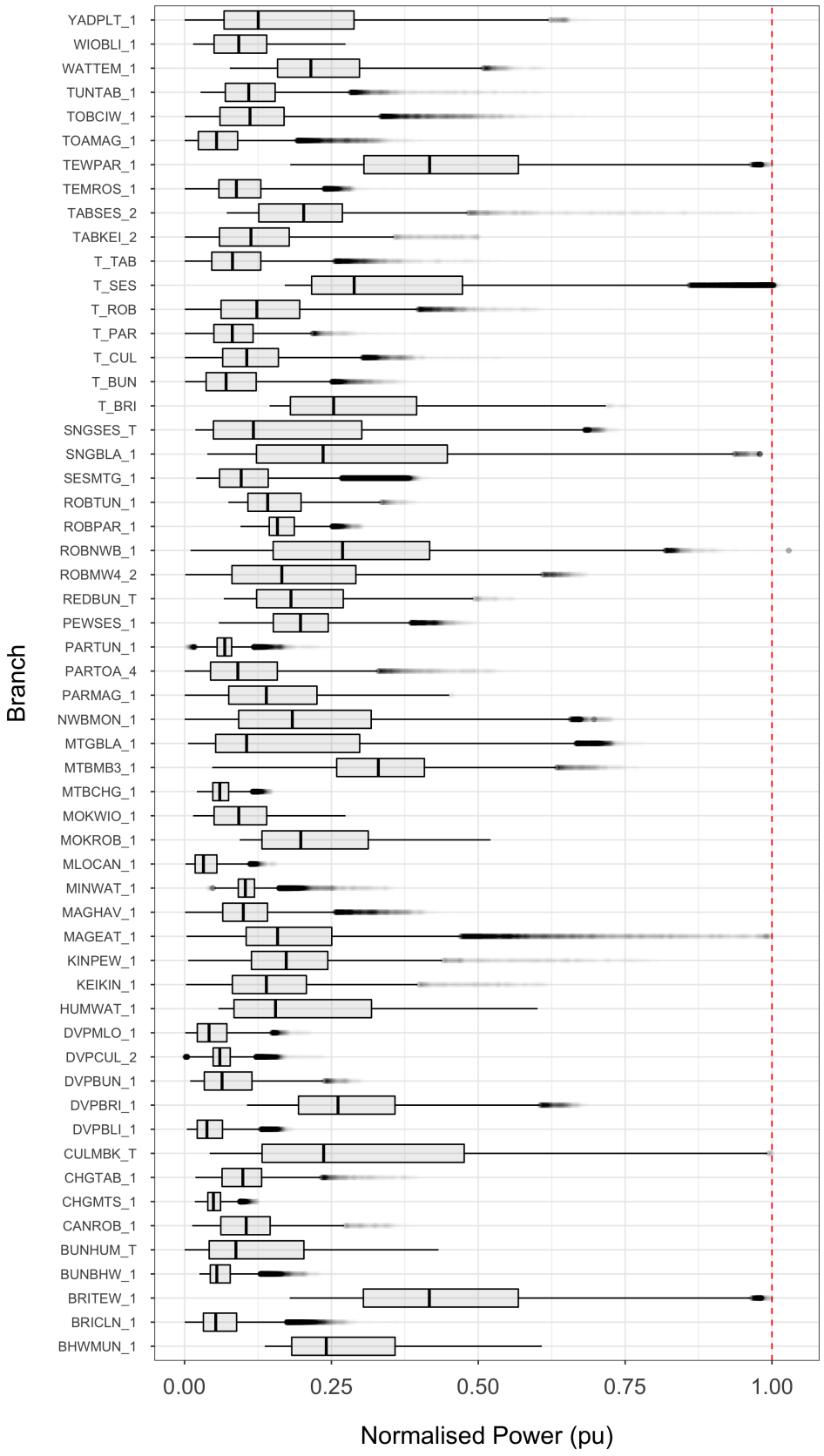


Figure 4.5: Diagnostic box-plots of the branch power flows (pu) for the AC model of the SA power network. There is a very small number of instances where the branch power flows exceed their capacity.

4.4 DC approximation

Whilst AC power flows are more realistic, no closed-form solution exists and therefore it is convenient to analyse DC power flows. The physical model of the AC model is used for the DC model. Most software packages can easily switch between AC and DC power flow calculations. However, the dynamic inputs from Section 4.2 contain transmission losses, whereas DC assumes a lossless network. Since the AC model underestimates demand, we proportionally increase load, as seen in Section 4.2.3, in order to satisfy the balancing equation.

In order to linearise the AC power flows a number of assumptions are made:

- Line resistance is negligible.
- Voltages are near 1 pu.
- Voltage angle differences are small.

In the subsequent sections, we address the validity of each assumption.

4.4.1 Line resistance

The first assumption that plays a major role in the accuracy of DC power flows, is the one of negligible line resistance. The X/R ratio is a condition number that quantifies how much the line reactance, X, dominates the line resistance, R. Larger ratios imply a higher accuracy of the DC approximation. The influence of resistance decreases when voltage increases, which means that only high voltage transmission lines withstand this condition.

In Table 4.2, we present typical line parameters for transmission lines of 132 kV and 275 kV, which are used in our AC model. The corresponding X/R ratios are displayed in Table 4.6.

	132 kV	275 kV
X/R	3.482	7.975

Table 4.6: The X/R ratio for the 132 kV and 275 kV transmission lines.

It can be seen that the X/R ratio for the 275 kV lines is sufficiently large to ensure small errors. However, the X/R ratio for the 132 kV lines potentially leads to errors around 5% [54], especially since the topology of the network includes very long 132 kV transmission lines which have a high resistance and experience large power losses.

4.4.2 Voltage variations

The second assumption of the DC approximation is that voltages are near 1 pu. However, it has been noted that the critical characteristic is a flat voltage profile rather than being near unity [54]. This is because nodal voltages operated above 1 pu experience decreased power losses, which is closer to the lossless assumption.

The individual voltage distributions for each bus are displayed in Figure 4.4 in Section 4.3. However, this does not quantify the magnitude of voltage deviations. We measure the voltage deviations by calculating their standard deviation

$$\sigma_V = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(V_i - \overline{V})^2}\,,$$

where $\overline{V}$ is the average voltage magnitude. We calculate the voltage deviation, σ_V, for each time point and plot the resulting distribution in Figure 4.6.

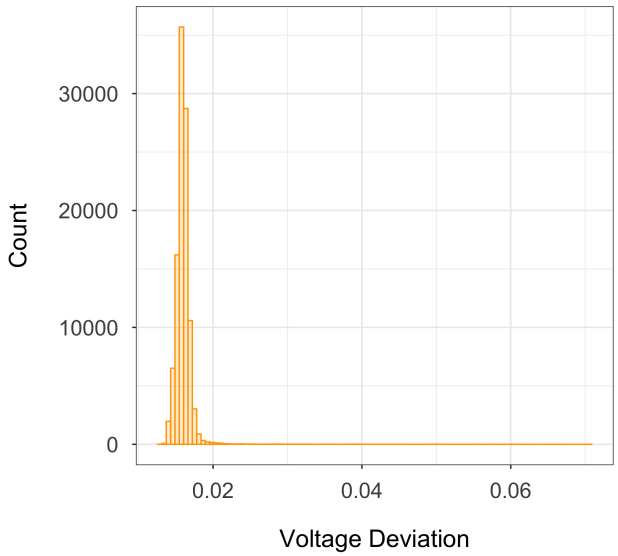


Figure 4.6: Distribution of the voltage deviation, σ_V, at each time point of the AC model. The vast majority of times have a voltage deviation less than 0.02.

It is evident that the location of the voltage deviation distribution is approximately 0.015, but the distribution has a long tail. Time points corresponding to voltage deviations that exceed 0.02 are not accurate. Once again, the X/R ratio for the 275 kV transmission lines is large enough to ensure the errors are less the 5%, whereas the X/R ratio for the 132 kV transmission lines corresponds to errors potentially above 5% [54].

4.4.3 Voltage angle differences

Denote by θ_k and θ_h the voltage angles of the *to* and *from* buses of the branch (k, h), respectively. Then, the line voltage angle difference is given by $\theta_k - \theta_h$. The assumption of small voltage angle differences results in the following approximations

$$\begin{aligned} \sin(\theta_k - \theta_h) &\approx \theta_k - \theta_h \,, \\ \cos(\theta_k - \theta_h) &\approx 1 \,. \end{aligned}$$

We calculate the line voltage angle differences for each time point of the SA network model, and assess the error of the sine and cosine approximations in Figure 4.7.

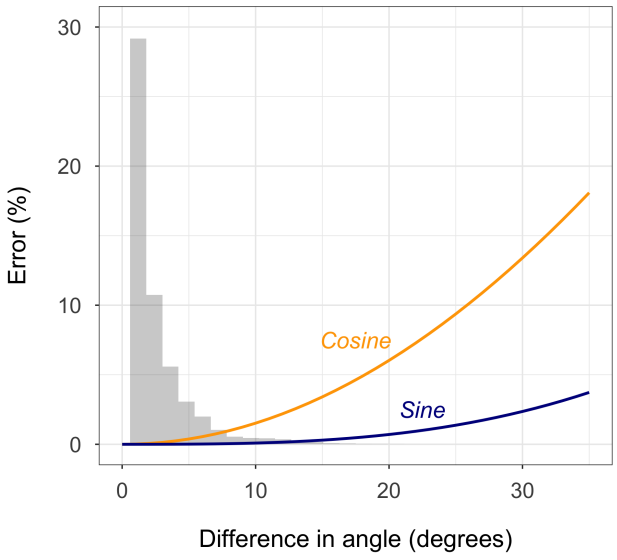


Figure 4.7: Error percentage of the *sine* and *cosine* approximations for the DC model. We include the (grey) distribution of the line voltage angle differences of the AC model. It can be seen that the location of the distribution coincides with small errors in the sine and cosine approximations.

The distribution of the line voltage angles is located where the *sine* and *cosine* approximations have minimal error. Therefore, we conclude that the assumption

of small voltage angle differences is reasonable and is the source of negligible error.

4.5 Power transmission distribution factors

The linear DC model calculates the matrix of PTDFs, H, which are the coefficients of the power injections, $\boldsymbol{p}$. Therefore, a change in the power injection will result in a change in the power flow along each line ℓ based on the PTDFs. This phenomenon may also be considered as the sensitivity of the power flow solution to changes in the power injections. This is a feature of the physical model described in Section 4.1, and is useful for diagnosing causes of congestion.

The sensitivity of the branch power flows is governed by the matrix of PTDFs, H. The sensitivity of branch ℓ to the power injection at bus k is given by

$$s_{\ell k} = |H_{\ell k}|\,,$$

where $s_{\ell k} = 1$ implies that a 1 MW increase in power injected at bus k induces a 1 MW change in branch ℓ. We visualise the sensitivity between each branch-bus pair as a heatmap in Figure 4.8. It is evident that the branch most sensitive to the power injection at the largest number of buses is the *Tailem Bend 275 - South East 275* transmission line. This transmission line is critical in exporting/importing power to and from Victoria.

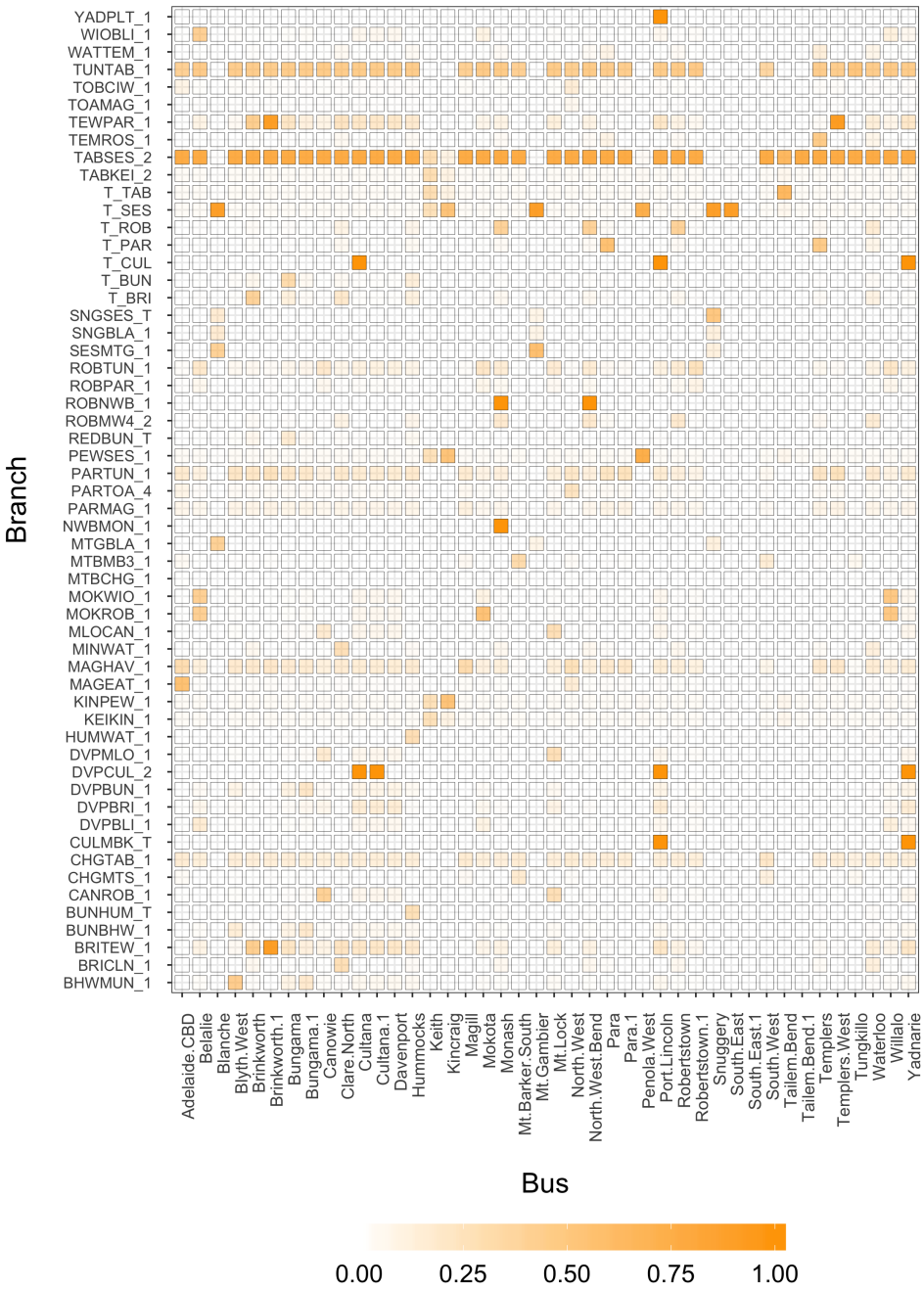


Figure 4.8: Visualisation of the PTDFs which represent the change in the real power flows due to a 1 MW increase in the power injections for the DC model. The most sensitive line is the 275 kV high voltage transmission line TABSES_2 which connects Tailem Bend and South East.

Chapter 5

Wind power simulation

The installed capacity of wind generation in SA is increased by incorporating new hypothetical wind farms into the *physical model* of the network developed in Section 4.1. The time series of generation for new wind farms is unknown and therefore requires simulation in order to be combined with the *dynamic inputs* from Section 4.2. In this chapter, we develop a wind power simulation methodology and then validate the method using historical wind generation data for SA.

5.1 Overview

In Chapter 4 we developed the *physical model* of the South Australian power network for 2015, which includes 17 operational wind farms with a combined capacity of 1,491 MW. We briefly introduce the wind farms in Table 5.1 and show their locations in Figure 5.1. We note that since 2015 the following wind farms have been commissioned: the 314 MW Hornsdale wind farm in 2017, the 119 MW Wilogoleche wind farm in 2018, and the first stage of the 212 MW Lincoln Gap wind farm in 2019.

The time series of wind generation for each wind farm is sourced from AEMO [45], and was introduced in Section 4.2.1. We plot a sample of the wind generation for the wind farms in the Mid-North (East) in Figure 5.2. It is evident that, despite variation between the time series of generation for each wind farm, there is also correlation. This is demonstrated between the 75th and 150th index, where the available wind resource obviously increases, and thus the power output of the wind farms also increases.

ID	Name	Zone	Capacity (MW)	Year
BLUFF1	The Bluff	Mid-North (East)	52	2012
CATHROCK	Cathedral Rocks	Eastern Eyre Peninsula	66	2005
CLEMGPWF	Clements Gap	Mid-North (West)	57	2010
CNUNDAWF	Canunda	South East	46	2005
HALLWF1	Hallett 1	Mid-North (East)	94	2007
HALLWF2	Hallett 2	Mid-North (East)	71	2009
LKBONNY1	Lake Bonney 1	South East	80	2005
LKBONNY2	Lake Bonney 2	South East	159	2008
LKBONNY3	Lake Bonney 3	South East	39	2010
MTMILLAR	Mount Millar	Eastern Eyre Peninsula	70	2006
NBHWF1	North Brown Hill	Mid-North	132	2008
SNOWNTH1	Snowtown North	Mid-North (West)	144	2014
SNOWSTH1	Snowtown South	Mid-North (West)	126	2014
SNOWTWN1	Snowtown	Mid-North (West)	99	2008
STARHLWF	Starfish Hill	Fleurieu Peninsula†	34	2003
WATERLWF	Waterloo	Mid-North (East)	131	2010
WPWF	Wattle Point	Yorke Peninsula	91	2005

† indicates a zone that is not officially recognised for development in the ISP.

Table 5.1: Overview of the wind farms in South Australia. The ID is the abbreviation used by AEMO; the zone is the renewable energy zone (REZ) where the wind farm is located according to AEMO's integration system plan (ISP) [47]; the capacity is the wind farms maximum power output; and the year refers to the year the wind farm started operation.

A *correlation matrix* is a matrix with elements consisting of pairwise correlation coefficients between sets of variables. This technique is predominantly used in visualisation in order to identify sets of correlated variables. When constructing the correlation matrix for time series data, we evaluate the *cross-correlation* between two time series. Denote by W_t^j the time series of generation for wind farm j measured in MW. Then, the time series $\{W_t^1\}_{t\in\mathbb{N}}, \{W_t^2\}_{t\in\mathbb{N}}, \ldots, \{W_t^{17}\}_{t\in\mathbb{N}}$ have a correlation matrix given by

$$[\,\rho\,]_{ij} = \rho_{w_i w_j}(0)\,,$$

where $\rho_{w_i w_j}(0)$ is *Spearman's rank correlation coefficient* [66] of $\{W_t^i\}_{t\in\mathbb{N}}$ and $\{W_t^j\}_{t\in\mathbb{N}}$. In Figure 5.3, we plot the correlation matrix of the time series of generation for all the wind farms.

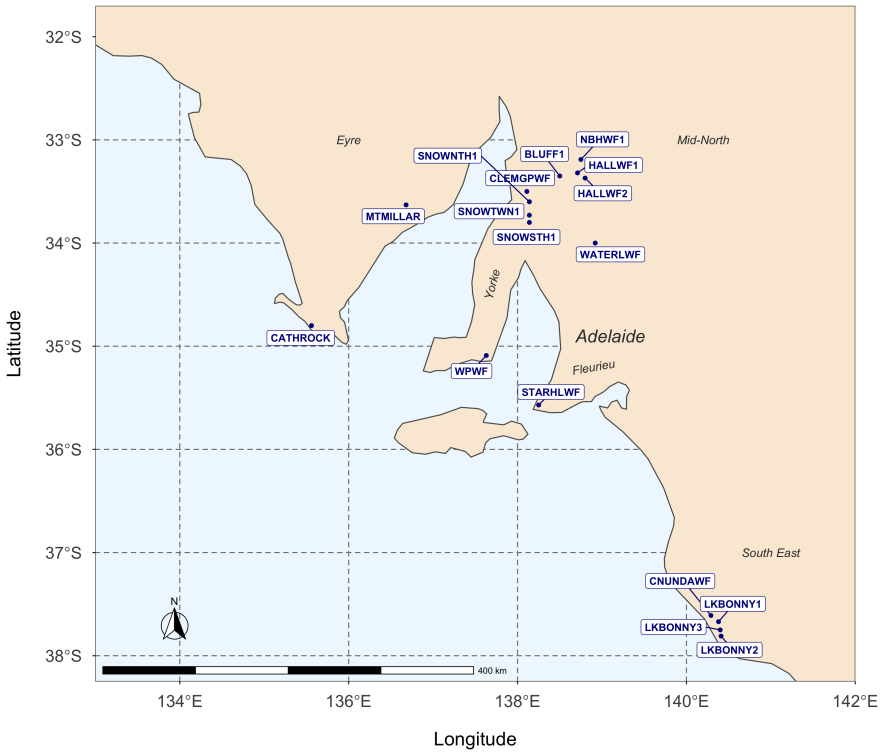


Figure 5.1: Map of the wind farm locations in South Australia. The Mid-North and South East contain the majority of installed wind capacity.

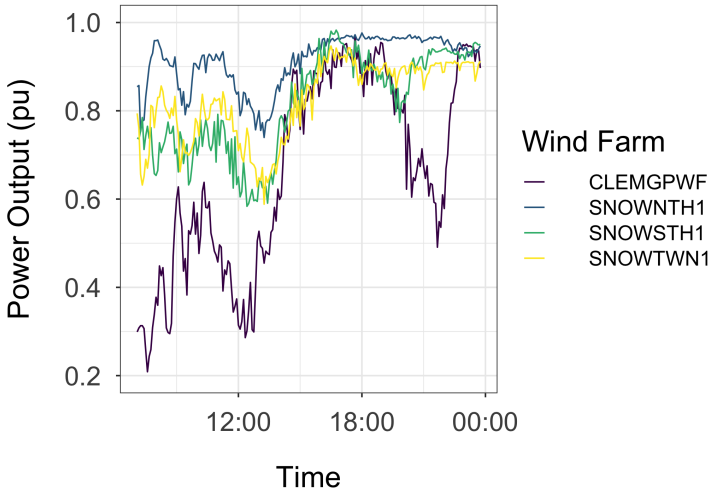


Figure 5.2: An example of the time series of generation for the wind farms in the Mid-North (East). The x-axis represents time and the y-axis is power output per-unit.

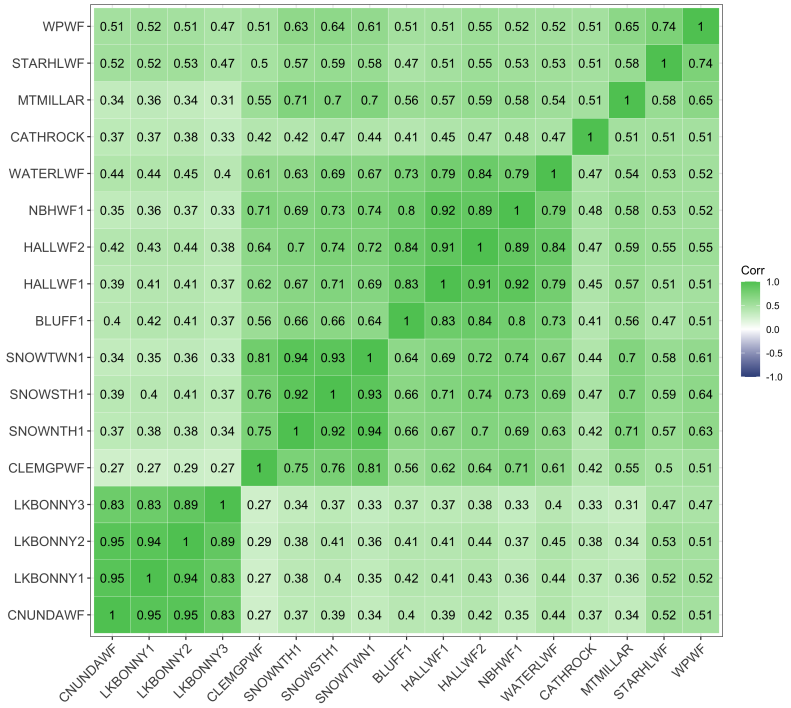


Figure 5.3: Correlation of wind generation in SA. The time series of generation for all wind farms are positively correlated.

It is evident that there are three distinct groups of highly correlated wind farms: South East, Mid-North (East) and Mid-North (West). A contributing factor to the high correlation presence is due to the well-known geographical dependency of the weather. Clustering of wind farms is common, as companies develop wind farms in areas where the wind resource is plentiful and there is access to transmission.

Wind generation is stochastic and variable on a 5-minute period. Whilst time series models seem reasonable, our analysis of the power spectrum showed that significant cycles explain only a small portion of the variance, see Appendix B. When the deterministic model explains only a small portion of the variance in the data, it leads to simulations that are too random such that there is a small correlation with existing wind farms. Since high correlation in wind generation is a significant driver for network congestion, we explore deterministic models such as *regression*.

Unlike time series models, regression models do not require de-trending since they exploit the correlations between the variables. The power output of a wind farm is bounded above by its generation capacity, and below by the power required to run essential equipment when there is insufficient wind resource. We apply a *logit*

transformation to the wind generation in Section 5.3.1 such that the model is fitted in an unbounded space. The transformed variables in Figure 5.4 demonstrate a linear relationship.

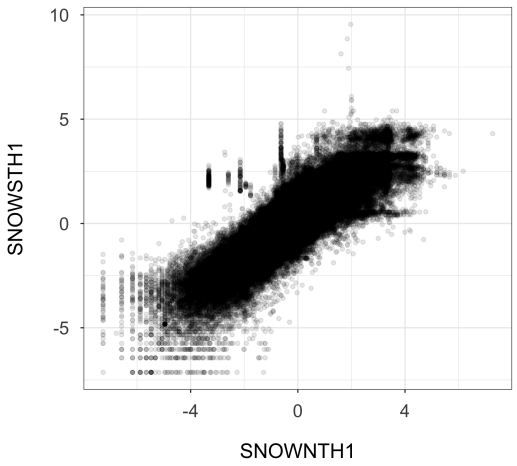


Figure 5.4: Scatterplot of the generation from SNOWSTH1 and SNOWNTH1 in the transformed *logit space*. The variables demonstrate a linear relationship.

As an example, we fit a simple linear regression that predicts the response variable SNOWNTH1 from the other wind farms in the Mid-North (West) given by

$$\text{SNOWNTH1} \sim \text{CLEMGPWF} + \text{SNOWSTH1} + \text{SNOWTWN1}\,.$$

From the summary of the regression, provided in Table 5.2, we see a high adjusted R-squared value. This suggests that a regression based model will be able to accurately predict the power output of one wind farm from the power output of nearby wind farms.

Term	Estimate	Standard Error	Test Statistic	p-value
(Intercept)	-1.189	0.081	-14.703	0.000
CLEMGPWF	-0.069	0.005	-14.611	0.000
SNOWSTH1	0.419	0.003	124.851	0.000
SNOWTWN1	0.781	0.005	171.727	0.000
Adjusted R-squared				0.8571

Table 5.2: Output from the linear regression that predicts the response wind farm SNOWNTH1. The high adjusted R-squared value demonstrates that the predictor wind farms explain a significant amount of the variation in the response.

We simulate the time series of generation for a new wind farm using historical wind generation of the existing wind farms and wind data. In Section 5.3 we fit a

number of linear regression models, which are used to predict the expected power output of the new wind farm, called the *mean prediction model*. However, this is a deterministic approach to modelling a stochastic variable. Hence, in Section 5.4, we simulate a stochastic deviation from the expected power output in order to obtain a realistic time series of wind generation, called the *error simulation model*. We assess the goodness-of-fit of the mean prediction model and the error simulation model by applying a *leave-one-out* technique such that one of the existing wind farms is treated as unknown. This emulates the prediction of a new wind farm, and allows us to produce diagnostic plots.

5.2 Data

The wind power simulation uses historical wind generation data from Section 4.2.1 and wind data from the Bureau of Meteorology (BOM) [41]. First, we introduce the wind data, and second, we describe the pre-processing steps performed on the data prior to fitting the model.

5.2.1 Meteorological data

There are a number of weather stations deployed throughout SA that are operated by the BOM and measure a range of atmospheric variables. In particular, they record the wind speed in m/s and the wind direction in degrees. The dataset contains wind speed and direction for the weather stations listed in Table 5.3 at a 1-minute resolution.

Station number	Name	Zone	Start year	Complete (%)
26021	Mt Gambier	South East	1941	99
21133	Snowtown (Rayville Park)	Mid-North	1998	98
21131	Clare High School	Mid-North	1993	99

Table 5.3: Basic weather station information. The station number is the BOM identification number, and the column Complete refers to the number of complete observations. The Mid-North weather stations are used for both East and West Mid-North wind farms.

The missing data at the 1-minute resolution is imputed by linear interpolation. Then the data are down-sampled to a 5-minute resolution to coincide with the 5-

minute dispatch. Since the data are down-sampled, the linear imputation will have a negligible effect, especially for wind direction.

Typically, the wind direction is a continuous variable that measures the angle in degrees, where 0 degrees is true North. For our purposes, we are not interested in measuring the complex relationship between direction and power generation; therefore, we factor the direction into a categorical variable. Denote by o_t^j the wind direction of weather station j at time t. The wind direction is factored such that

$$\Theta_t^j = \begin{cases} NNE & 0 \leq o_t^j < 45\,, \\ ENE & 45 \leq o_t^j < 90\,, \\ ESE & 90 \leq o_t^j < 135\,, \\ SSE & 135 \leq o_t^j < 180\,, \\ SSW & 180 \leq o_t^j < 225\,, \\ WSW & 225 \leq o_t^j < 270\,, \\ WNW & 270 \leq o_t^j < 315\,, \\ NNW & 315 \leq o_t^j < 360\,, \end{cases}$$

where Θ_t^j represents the categorical wind direction of weather station j at time t.

5.2.2 Data processing

Prior to fitting the mean prediction model, we perform pre-processing steps on the wind generation data. First, we normalise the power generation by the generators capacity to eliminate scaling effects, which is important for Section 5.3.1. Second, we remove all observations that contain non-positive power generation values, to ensure that $0 < W_t^j < 1$, which is a required condition to perform the logit transformation. The non-positive values arise when wind speeds are below the turbine cut-in speed, which is the minimum required speed before the turbines produce power, and therefore the wind farm consumes power to run essential equipment. Removing non-positive values for estimation is reasonable because network congestion occurs only during windy periods when wind farms are operating at high power outputs.

Additionally, the data contains periods of maintenance that effect the wind generation profiles as seen in Figure 5.5. Such behaviour in the data biases the parameter estimation. We filter out the periods of maintenance using the method described in Appendix C.

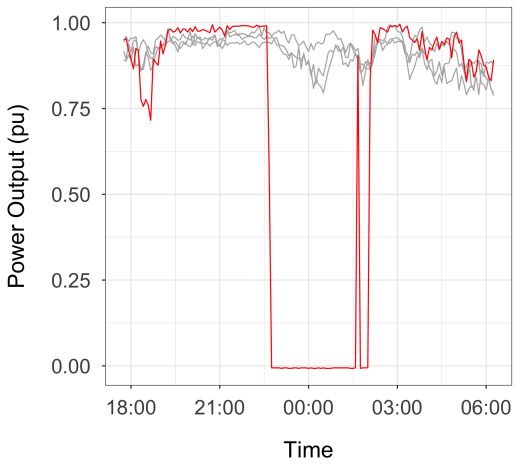


Figure 5.5: An example of maintenance affecting the power output of a wind farm in the South East. The x-axis represents time and the y-axis is the power output per-unit. The (red) highlighted time series corresponds to a wind farm that experiences a sudden loss of power for approximately 3 hours. This is inconsistent with the available wind resource in the South East, indicated by the (grey) time series of generation of the other nearby wind farms.

5.3 Mean prediction model

Consider the new wind farm to be located in a zone where there are currently k operational wind farms with corresponding power generation $W_t^1, W_t^2, \ldots, W_t^k$ measured in MW, and m weather stations, where $\Phi_t^1, \Phi_t^2, \ldots, \Phi_t^m$ and $\Theta_t^1, \Theta_t^2, \ldots, \Theta_t^m$ represent the corresponding wind speeds (m/s) and the categorical wind directions, respectively. Initially, we propose a method of regression called the *average regression model*, and then improve this model in the *corrected average regression model*.

5.3.1 Average regression model

The average regression model is designed to estimate the average effect that an existing wind farm has on the the other wind farms. Then, the average effect of all existing wind farms can be used to predict the power output of a new wind farm.

For each wind farm $j = 1, \ldots, k$, we regress the power generation W_t^{j*} on all other wind farms, where * indicates a variable in the logit transformed space such

that

$$W_t^{j*} = \text{logit}(W_t^j) = \log\left(\frac{W_t^j}{1 - W_t^j}\right).$$

The logit transformation is used to allow for the fact that the normalised power generation is constrained to lie in the range $0 < W_t^j < 1$. This results in a total of k models, where Model j is given by

$$\begin{aligned} W_t^{j*} = \alpha_{1,j}W_t^{1*} + \cdots + \alpha_{j-1,j}W_t^{j-1*} + \alpha_{j+1,j}W_t^{j+1*} + \cdots + \alpha_{k,j}W_t^{k*} + \\ \beta_{1,j}\Phi_t^1 + \cdots + \beta_{m,j}\Phi_t^m + \gamma_{1,j}\Theta_t^1 + \cdots + \gamma_{m,j}\Theta_t^m, \end{aligned} \tag{5.1}$$

where $\alpha_{i,j}$ is the regression coefficient of wind farm i in Model j, $\beta_{h,j}$ and $\gamma_{h,j}$ are the regression coefficients of wind speed and wind direction, respectively, of weather station h in Model j. Wind direction is a categorical variable measured on eight levels, refer to Section 5.2.1. The regression model in Equation (5.1) estimates a coefficient for each level such that

$$\gamma_{h,j} = \mathbb{1}_{NNE}\xi_{NNE,j} + \cdots + \mathbb{1}_{NNW}\xi_{NNW,j},$$

where $\mathbb{1}$ is the indicator function and $\xi_{\ell,j}$ is the coefficient for Level ℓ in Model j. We use the $\gamma_{h,j}$ variables in (5.1) to simplify the notation. Then, the fitted values of the average regression model are given by

$$\eta_t^* = \overline{\alpha}_1 W_t^{1*} + \cdots + \overline{\alpha}_k W_t^{k*} + \overline{\beta}_1\Phi_t^1 + \cdots + \overline{\beta}_m\Phi_t^m + \overline{\gamma}_1\Theta_1 + \cdots + \overline{\gamma}_m\Theta_m, \tag{5.2}$$

where the average regression coefficients are given by

$$\begin{aligned} \overline{\alpha}_i &= \frac{1}{k}\sum_{j=1}^{k}\alpha_{i,j}, \\ \overline{\beta}_h &= \frac{1}{k}\sum_{j=1}^{k}\beta_{h,j}, \\ \overline{\gamma}_h &= \mathbb{1}_{NNE}\overline{\xi}_{NNE} + \cdots + \mathbb{1}_{NNW}\overline{\xi}_{NNW}, \end{aligned}$$

where $\overline{\xi}_\ell = \frac{1}{k}\sum_{j=1}^{k}\xi_{\ell,j}$.

The average regression model is deterministic, and therefore the same inputs produce the same fitted values, η_t^*. Applying a *leave-one-out* technique, we treat SNOWNTH1 as unknown, and fit the average regression model using the Snowtown and Clare High school weather stations, and the other wind farms within the South East: CLEMGPWF, SNOWSTH1 and SNOWTWN1. We plot the residuals versus fitted values of the average regression model in Figure 5.6. We repeat the leave-one-out process for the other 12 wind farms from the South East and Mid-North, and provide their diagnostic plots in Appendix D Figure D.1.

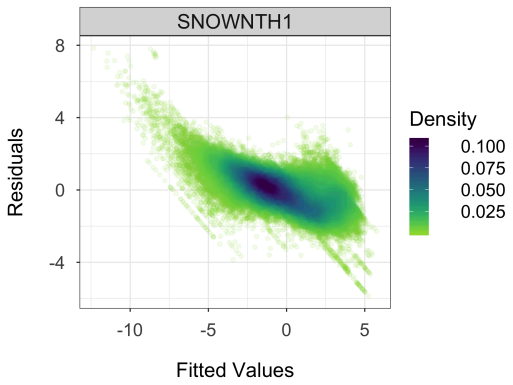


Figure 5.6: Residual versus fitted values of the average regression model that predicts SNOWNTH1. The x-axis is the fitted values, η_t^*, and the y-axis is the difference between the fitted values and the actual values. A linear trend-line is evident in the residuals.

The residuals in Figure 5.6 contain a linear trend, which is induced by the estimation procedure. The average regression coefficients are estimated using $k-1$ wind farm predictors in Equation (5.1), as one wind farm is the response variable, whereas the fitted values are calculated using all k wind farms in Equation 5.2. The extra predictor causes the fitted values to be biased, and thus resulting in a linear trend in the residuals. We propose a post-corrective method to remove the linear trend, called the *corrected average regression model*.

5.3.2 Corrected average regression model

In order to correct the bias induced by the estimation of the average regression model, we estimate the linear trend observed in the residuals. First, we calculate the residuals from the individual regression models from Equation (5.1) such that

$$e_t^{j*} = W_t^{j*} - \eta_t^* \,,$$

where e_t^{j*} are the residual values for wind farm $j = 1, \ldots, k$. The linear relationship between the residuals and the fitted values is given by

$$e_t^{j*} = s_j \eta_t^* + r_j \,,$$

where s_j and r_j represent the slope and intercept of Model j, respectively. Then, the fitted values of the corrected average regression are given by

$$\widehat{\eta}_t^* = \eta_t^* + \overline{s}\eta_t^* + \overline{r} \,, \tag{5.3}$$

where $\overline{s}$ and $\overline{r}$ represent the mean slope and intercept for the k wind farms, respectively. We plot the residuals versus fitted values of the corrected average regression model in Figure 5.7, when SNOWNTH1 is treated as unknown. Similarly, the supplementary diagnostics of the other 12 wind farms are provided later in Figure D.2.

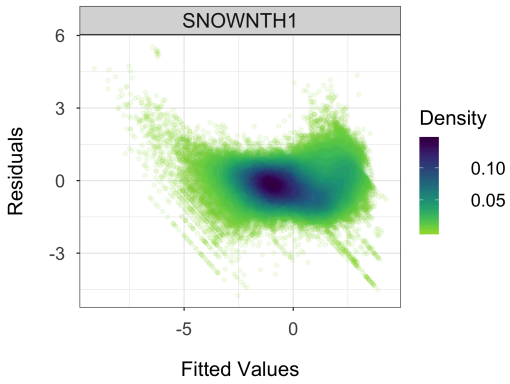


Figure 5.7: Residual versus fitted values of the corrected average regression model that predicts SNOWNTH1. The x-axis is the fitted values, $\widehat{\eta}_t^*$, and the y-axis is the difference between the fitted values and the actual values. The Spearman's rank correlation coefficient of -0.06 is statistically significant and suggests a small curvilinear trend.

Figure 5.7 demonstrates that the fitted values of the corrected average regression significantly reduce the linear trend present in the average regression from Figure 5.6. The average Spearman's rank correlation coefficient of -0.08 for all wind farms, which includes Figures 5.7 and D.2, is statistically significant and suggests a small curvilinear trend. Furthermore, the conditional density of the residuals given the fitted values has a uni-modal structure consistent with a bivariate normal. However, there still exists negative slope excursions near extreme values corresponding to the generation limits. These are attributed to maintenance, where at least one wind farm is not operating normally, and truncation due to capacity limits. This suggests that our filtering of maintenance in Section 5.2.2 could be improved. However, this could be expected to have a minimal effect on the parameter estimation due to the large dataset.

The fitted values of the corrected average regression are referred to as the *mean predicted values* and are deterministic. Therefore, we develop a stochastic model using the residuals of the *corrected average regression model* in order to simulate realistic stochastic wind generation.

5.4 Error simulation model

Local variations in atmospheric conditions result in wind generation being stochastic. This means that wind farms will fluctuate around the mean prediction level with a specified standard deviation. We develop a time series model that simulates the error around the mean prediction such that the time series of generation for a new wind farms is stochastic.

5.4.1 Time series model

The difference between the mean prediction values from the corrected average regression model, $\widehat{\eta}_t^*$, and the actual values of each wind farm $j = 1, \ldots, k$ is given by

$$\widehat{e}_t^{j*} = W_t^{j*} - \widehat{\eta}_t^* \, .$$

We know that the residuals of the corrected average regression model, $\widehat{e}_t^{j*}$, contain a small linear trend which will induce auto-correlation. We model the residuals, $\widehat{e}_t^{j*}$, using an ARMA(2,1) model such that

$$\widehat{e}_t^{j*} = a_{1j}\widehat{e}_{t-1}^{j*} + a_{2j}\widehat{e}_{t-2}^{j*} + \varepsilon_t + m_{1j}\varepsilon_{t-1} \, ,$$

where a_{1j} and a_{2j} are the autoregressive coefficients, m_{1j} is the moving average coefficient, and $\{\varepsilon_t\}_{t\in\mathbb{N}}$ is the white noise process. This gives us k ARMA models for the residuals, one for each existing wind farm. As the actual residuals for a new wind farm cannot be obtained, we estimate the parameters of the average ARMA(2,1) model such that

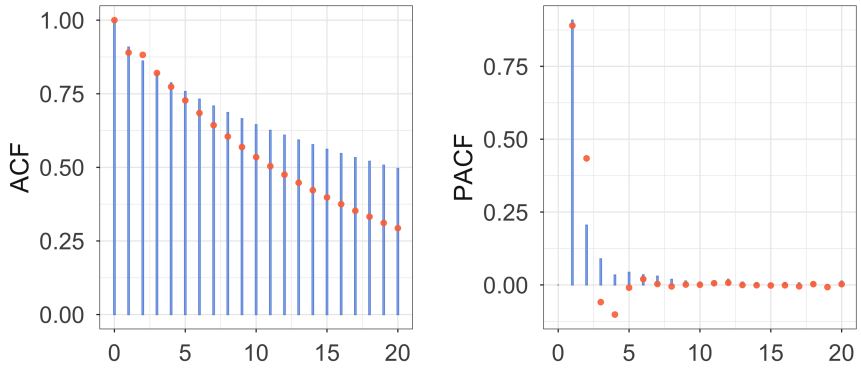
$$\begin{aligned} \overline{a}_1 &= \frac{1}{k}\sum_{j=1}^{k} a_{1j} \, , \\ \overline{a}_2 &= \frac{1}{k}\sum_{j=1}^{k} a_{2j} \, , \\ \overline{m}_1 &= \frac{1}{k}\sum_{j=1}^{k} m_{1j} \, . \end{aligned}$$

We simulate the average ARMA(2,1) to obtain the *simulated residuals*, ζ_t^*. Diagnostic plots of the auto-correlation and partial auto-correlation of the simulated residuals are displayed in Figure 5.8.

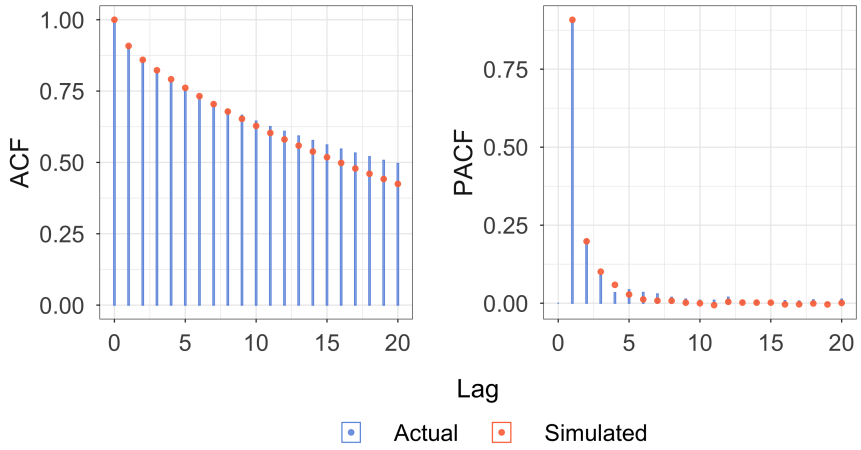
Figure 5.8 demonstrates that the auto-correlation and partial auto-correlation of the simulations of the ARMA(2,1) successfully model the strong auto-regressive

nature of the residuals. We found that simpler models, such as the AR(2), did not accurately represent the partial auto-correlation.

The time series model captures the auto-correlation structure but assumes a fixed Gaussian conditional density between the residuals and the fitted values. In practice, we observe that the conditional density near capacity has a smaller standard deviation and a smaller mean, refer to Figure 5.7. Therefore, we use quantile regression to transform the simulated residuals, ζ_t^*.



(a) Simulations from the AR(2) model.



(b) Simulations from the ARMA(2,1) model.

Figure 5.8: The autocorrelation (left) and the partial autocorrelation (right) of the residuals and the simulated residuals. The simulated residuals are simulations from a time series model fitted to the residual values of the corrected average regression model that predicts SNOWNTH1. (a) Simulations from the AR(2) model, and (b) simulations from the ARMA(2,1) model with supplementary plots in Figures D.3 and D.4. The more complex ARMA model captures the partial auto-correlation structure.

5.4.2 Quantile regression transformation

The quantile regression transformation fits two models: an *upper* and a *lower* quantile. The two quantiles outline the shape of the residuals in Figure 5.7. We then transform the ARMA time series simulation to the corresponding residual quantiles. This extends the parametric time series simulation to a non-parametric approach.

Denote by $Q_{\widehat{e}_t^{j*}|\widehat{\eta}_t^*}(\tau)$ the τth quantile function for the conditional distribution of residuals, $\widehat{e}_t^{j*}$, given the mean predicted values, $\widehat{\eta}_t^*$. The quantile function is modelled as a natural cubic spline on K knots such that

$$Q_{\widehat{e}_t^{j*}|\widehat{\eta}_t^*}(\tau) = \sum_{n=1}^{K+2} \kappa_n b_n(\widehat{\eta}_t^*)\,,$$

where $b_n(x)$ is the nth basis functions of the natural cubic spline, and κ_n is the corresponding quantile regression coefficient. Therefore, this provides k estimates of the residual conditional distribution, one for each existing wind farm. We average the quantile functions in order to obtain the averaged quantile function

$$\overline{Q}_{\widehat{e}_t^*|\widehat{\eta}_t^*}(\tau) = \frac{1}{k}\sum_{j=1}^{k} Q_{\widehat{e}_t^{j*}|\widehat{\eta}_t^*}(\tau)\,, \tag{5.4}$$

which evaluates the mean residual conditional distribution. This averaged quantile function is used to transform the simulated values from the time series model to fit the quantiles of the residuals. Since we average over the quantile functions, the choice of knots becomes less significant. Additionally, the fitted values of different wind farms in the logit space vary, and therefore the knots are given by the quantile function of the fitted values, $Q_{\widehat{\eta}_t^*}(\tau)$, evaluated at $\tau = 0.01, 0.1, 0.2, 0.4, 0.6, 0.8, 0.9, 0.99$.

As the max-min data range is unstable, we fit the 5th and 95th percentile for the lower and upper quantiles, respectively. Given a time $t > 0$ such that the corresponding mean predicted value $\widehat{\eta}_t^* = \eta$ and the simulated residual $\zeta_t^* = \zeta$, we evaluate the lower and upper average quantiles $\overline{Q}_{\widehat{e}_t^*|\widehat{\eta}_t^*=\eta}(0.05)$ and $\overline{Q}_{\widehat{e}_t^*|\widehat{\eta}_t^*=\eta}(0.95)$, respectively. The simulated residuals are linearly transformed such that

$$\psi_t^*(\zeta_t^* = \zeta|\widehat{\eta}_t^* = \eta) = s_\eta \zeta + r_\eta\,, \tag{5.5}$$

are called the *transformed simulated residuals*. Here, the slope s_η and intercept r_η are solved simultaneously using the linear system given by

$$\begin{aligned}
\overline{Q}_{\widehat{e}_t^*|\widehat{\eta}_t^*}(0.95) &= s_\eta Q_{\zeta_t^*}(0.95) + r_\eta\,,\\
\overline{Q}_{\widehat{e}_t^*|\widehat{\eta}_t^*}(0.05) &= s_\eta Q_{\zeta_t^*}(0.95) + r_\eta\,,
\end{aligned}$$

where $Q_{\zeta_t^*}(0.95)$ is the quantile function for the simulated values of the average ARMA(2,1) process evaluated at $\tau = 0.95$. This technique approximates the distribution of the residuals for each mean predicted value with a normal distribution fitted to the 5th and 95th percentiles. In Figure 5.9, we compare the lower, median and upper quantiles of the transformed simulated residuals to the actual residual quantiles in the transformed space. A negative logit difference implies that the transformed simulated residuals *underestimate* the actual residual quantile, and a positive logit difference implies that the transformed simulated residuals *overestimate* the actual residual quantile.

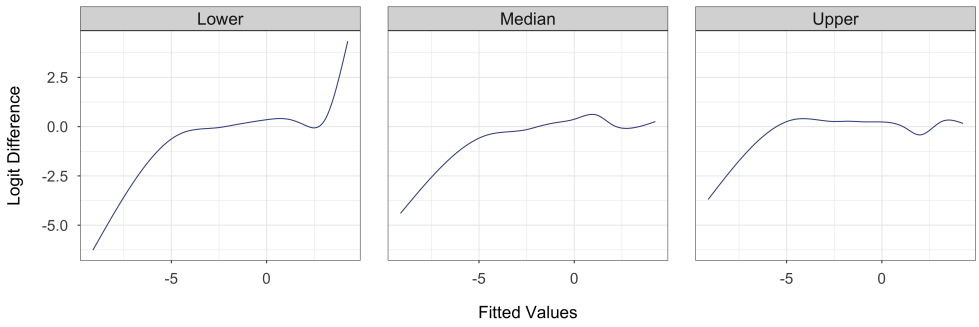


Figure 5.9: The x-axis is the fitted values, $\widehat{\eta}_t^*$, of the corrected average regression model fitted to SNOWNTH1. The y-axis is the logit difference between the averaged quantile function, $Q_{\widehat{e}_t^*|\widehat{\eta}_t^*}(\tau)$, and the actual residual quantiles evaluated at $\tau = 0.05, 0.5, 0.95$. Refer to Figure D.5 for supplementary plots.

It is evident that the method underestimates all three quantiles corresponding to low wind generation values. At high wind generation values, the median and the upper quantiles are reasonable estimates, whereas the lower quantile overestimates the lower quantile of the residuals. It is important to note that the method uses the residuals of existing wind farms in order to simulate the residuals for the new wind farm. If the response variable was known, the method of quantile transformation would likely have greater accuracy.

The transformed simulated residuals $\psi_t^*(\zeta_t^*|\widehat{\eta}_t^*)$ are combined with the mean predicted values $\widehat{\eta}_t^*$ such that

$$W_t^{\text{new}*} = \widehat{\eta}_t^* + \psi_t^*(\zeta_t^*|\widehat{\eta}_t^*)\,,$$

which implies

$$W_t^{\text{new}} = \frac{\exp(\widehat{\eta}_t^* + \psi_t^*(\zeta_t^*|\widehat{\eta}_t^*))}{1 + \exp(\widehat{\eta}_t^* + \psi_t^*(\zeta_t^*|\widehat{\eta}_t^*))}\,, \tag{5.6}$$

where W_t^{new} is the simulated power output in MW of a new wind farm at time t. We validate our wind power simulation methodology in the next section.

5.5 Validation

The model diagnostics provide a high level of detail regarding the ability of the simulation method to predict the existing wind farms. However, network congestion occurs when the total amount of wind generation in a particular zone is significant. This is due to generators in the same zone sharing transmission capacity, which leads to network elements being heavily loaded and possibly operating outside their stability limits. Therefore, we validate the wind power simulation model by comparing simulated generation values with the total wind generation for each zone as opposed to individual wind farms. We simulate 50 new wind farms for each zone using our wind power simulation methodology, and calculate the *correlation* and *Probability of Exceedance* (POE). Correlation refers to the strength of the relationship between the relative movements of two time series, and POE measures the proportion of time a time series spends above a given value.

Simulations need to be correlated with the existing wind farms such that high wind generation levels are obtained during windy periods. Figure 5.10 demonstrates that the new wind farms are highly correlated with existing generation. The spread of correlation for the South East is very small due to its high within-zone correlation. On the other hand, wind farms in the Mid-North (West) are more spatially distributed resulting in a slightly lower location for the correlation distribution and a larger variance.

Congestion occurs when wind farms are operating at high power outputs. Therefore, we assess the 75% POE which corresponds to the proportion of time the power output of a wind farm is above 75% of its capacity. It is evident in Figure 5.11 that the POE for the observed total generation is not contained within the POE distribution of the simulations. The wind power simulation predicts the 75% POE within less than 2% on average.

Despite a relatively small inflation in the POE, the simulation models appear to be producing accurate realistic stochastic wind generation and closely following the total wind generation profiles in Figure 5.12. Therefore, we conclude that our wind power simulation methodology is reasonable and that the errors are small.

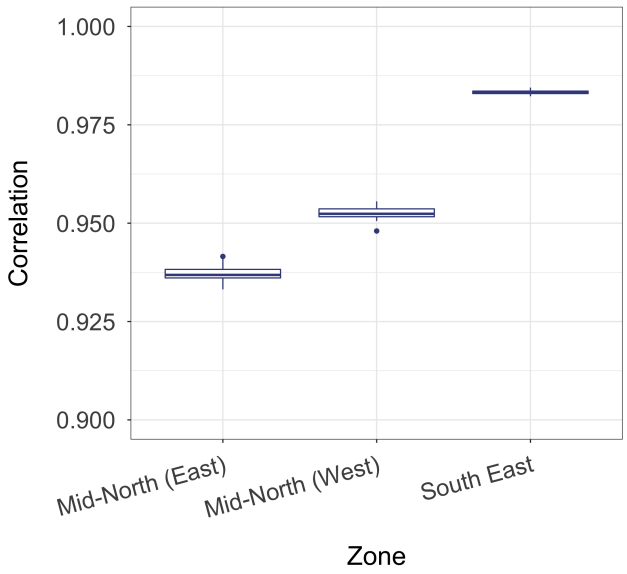


Figure 5.10: Distribution of the correlation between simulated and existing total wind generation for each zone. Simulations are highly correlated with existing wind generation.

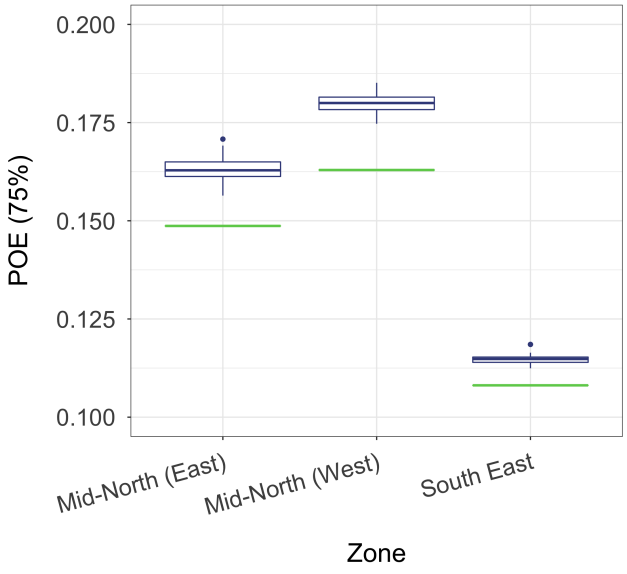


Figure 5.11: Distribution of the 75% probability of exceedance for each zone. The green line represents the actual POE of the existing total wind generation for each zone. Simulations marginally overestimate the POE.

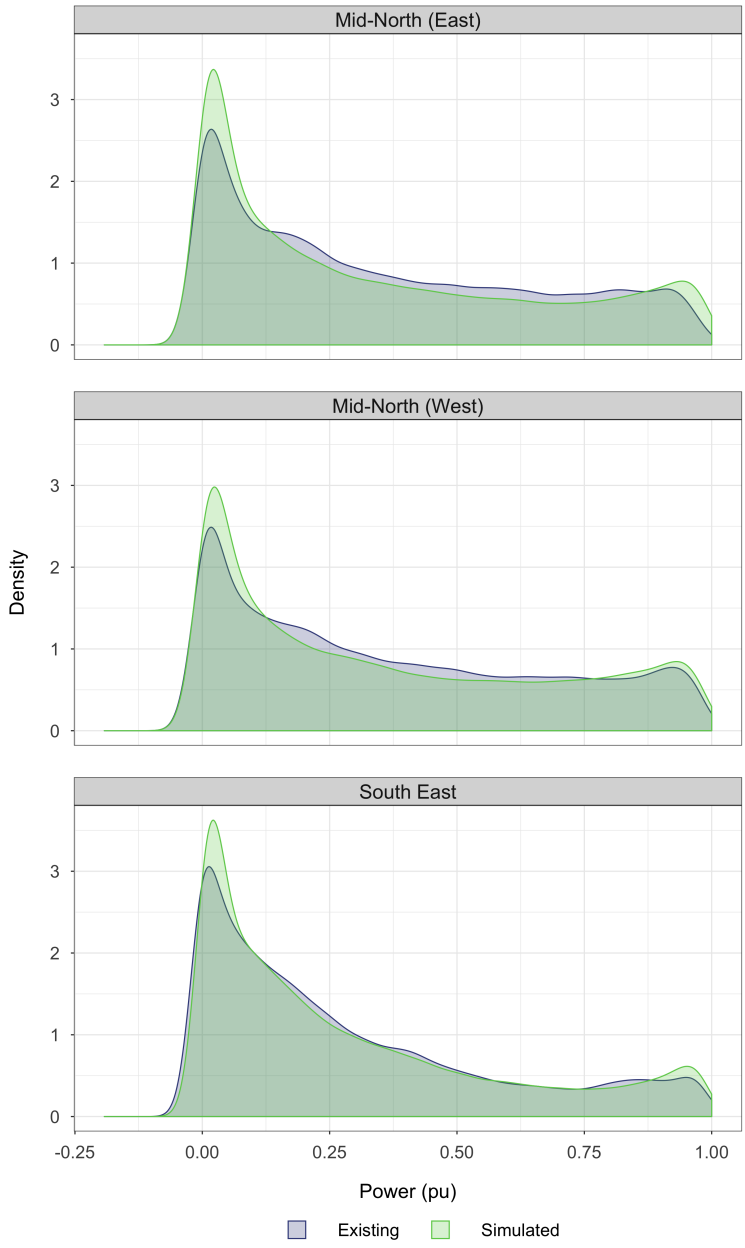


Figure 5.12: Distribution of total wind generation and simulated wind generation for each zone.

Chapter 6

Congestion and wind curtailment

A number of network congestion and wind curtailment case studies have been performed [7, 29, 33, 34], but none have been applied to the Australian power network. Whilst wind penetration in the NEM is relatively small, wind penetration in SA is almost 40% [52]. We perform a first analysis into quantifying wind curtailment that results from connecting additional wind capacity to the SA power network.

In Section 6.1, we develop two *short-term re-dispatch* models in order to estimate the amount of wind curtailment. In Section 6.2, we develop a *long-term congestion management* model that quantifies the impact of connecting new wind farms on network congestion and wind curtailment. Then, in Section 6.3, we present results of our methodology applied to the SA power network. We conclude this chapter with a discussion of the limitations of our model and future work.

6.1 Short-term re-dispatch

There are two main re-dispatch methods: targeted and global redistribution. Targeted methods will nominate a single wind farm to curtail [24], whereas global redistribution is more flexible, as it manages the output of all generators. Typically, global redistribution is solved using optimisation [12, 35, 63].

We formulate two optimisation problems: a linear and a quadratic. The linear optimisation maximises the amount of wind generation within the re-dispatch solution.

This is time independent and therefore depends only on the current demand profile. On the other hand, the quadratic optimisation minimises the change from the most economic dispatch solution. This helps to satisfy more technical constraints such as generator ramping rates.

Consider a set of power injections, $\boldsymbol{p} \in \mathbb{R}^n$, which produce a set of power flows, $\boldsymbol{f} \in \mathbb{R}^m$, where there exists at least one branch, ℓ, that is congested such that $|f_\ell| \geq 1$. Then, the re-dispatch model calculates a new set of power injections, $\boldsymbol{p}' \in \mathbb{R}^n$, such that no branch is congested so that $|\boldsymbol{f}'| \leq \mathbf{1}$, where $\boldsymbol{f}' \in \mathbb{R}^m$ is the power flows that result from re-dispatch and $\mathbf{1}$ is a column vector with all entries equal to one. We separate the power injections into two components: a variable component, $\hat{\boldsymbol{p}} \in \mathbb{R}^{n_1+1}$, which consists of n_1 generator buses and the slack bus, and a fixed component, $\bar{\boldsymbol{p}} \in \mathbb{R}^{n_2}$, consisting of n_2 load buses, such that $n = n_1+n_2+1$. We note that the fixed component, $\bar{\boldsymbol{p}}$, remains constant after re-dispatch and therefore $\bar{\boldsymbol{p}}' = \bar{\boldsymbol{p}}$. We assume that no load is lost and therefore the load buses do not change. Hence, we optimise over the variable component, which reduces the problem dimension and the number of constraints.

6.1.1 Maximise wind generation

The linear optimisation performs generation re-dispatch in order to maximise wind generation whilst satisfying the network constraints. The objective function is given by

$$\max_{\hat{p}'_k} \sum_{k=1}^{n} \mathbb{1}_{k\in\mathcal{I}_W} \hat{p}'_k \,, \tag{6.1}$$

where $\mathbb{1}_{k\in\mathcal{I}_W}$ is the indicator for a bus connected to a wind farm. Maximising wind generation is a heuristic, which attempts to replicate other optimisation objectives such as minimising operational costs. Renewable energy is the cheapest fuel source and will displace conventional generation in power markets. This is also equivalent to minimising wind curtailment in the re-dispatched solution.

We enforce that the re-dispatch solution must satisfy transmission constraints by assuming a DC approximation. Equation (2.4) states the linear DC power flows $\boldsymbol{f} = H\boldsymbol{p}$, where $H \in \mathbb{R}^{m\times n}$ is the matrix of PTDFs. Denote by $\hat{H} \in \mathbb{R}^{m\times(n_1+1)}$ and $\bar{H} \in \mathbb{R}^{m\times n_2}$ the matrix of PTDFs restricted to the variable and fixed buses, respectively. Then, the transmission constraints, $|\boldsymbol{f}'| \leq \mathbf{1}$, imposed by the re-dispatch, $\boldsymbol{p}'$, are given by

$$|H\boldsymbol{p}'| \leq \mathbf{1} \,,$$

which implies

$$\hat{H}\hat{\boldsymbol{p}}' \leq \mathbf{1} - \bar{H}\bar{\boldsymbol{p}}\,, \tag{6.2}$$

$$-\hat{H}\hat{\boldsymbol{p}}' \leq \mathbf{1} + \bar{H}\bar{\boldsymbol{p}}\,. \tag{6.3}$$

Additionally, the DC approximation assumes a lossless network which results in the power balancing equation

$$\mathbf{1}^\top \boldsymbol{p}' = 0\,. \tag{6.4}$$

A load bus does not have generation capacity, whereas a generator bus may have a load connected. Therefore, the new power injections, $\hat{\boldsymbol{p}}'$, are bounded above by the difference between the available generation, $\hat{\boldsymbol{g}}^{avail} \in \mathbb{R}^{n_1+1}$, and the current load, $\hat{\boldsymbol{d}} \geq \mathbf{0} \in \mathbb{R}^{n_1+1}$. We assume no curtailment is present in the original power injections, $\hat{\boldsymbol{p}}$, and therefore wind farms cannot increase generation in the re-dispatch solution, $\hat{\boldsymbol{p}}'$. On the other hand, conventional generators can increase up to their maximum capacity as they are not limited by available resource. The available generation at bus k is given by

$$\hat{g}_k^{avail} = \begin{cases} \hat{p}_k & k \in \mathcal{I}_W\,, \\ P_k^{\max} & \text{otherwise}\,, \end{cases}$$

where $P_k^{\max}$ is the maximum generation capacity at bus k. The load must remain fixed in the re-dispatch and therefore the upper bound is given by

$$\hat{\boldsymbol{p}}' \leq \hat{\boldsymbol{g}}^{avail} - \hat{\boldsymbol{d}}\,. \tag{6.5}$$

Similarly, the fixed load also implies that the re-dispatch, $\hat{\boldsymbol{p}}'$, is bounded below by the demand, $\hat{\boldsymbol{d}}$. However, as we notice in Chapter 5, wind farms consume power for essential equipment when there is no wind. Therefore, available generation at a bus may be negative and have no load connected. In order to ensure the problem is feasible, we take the minimum of demand and available generation such that

$$\hat{\boldsymbol{p}}' \geq \min(-\hat{\boldsymbol{d}}, \hat{\boldsymbol{g}}^{avail})\,. \tag{6.6}$$

Combining Equations (6.1)-(6.6), the linear optimisation problem is given by

$$\begin{aligned} \max_{\hat{p}'_k} \quad & \sum_{k=1}^{n} \mathbb{1}_{k\in\mathcal{I}_W} \hat{p}'_k \\ \text{s.t.} \quad & \hat{H}\hat{\boldsymbol{p}}' \leq \mathbf{1} - \bar{H}\bar{\boldsymbol{p}}\,, \\ & -\hat{H}\hat{\boldsymbol{p}}' \leq \mathbf{1} + \bar{H}\bar{\boldsymbol{p}}\,, \\ & \mathbf{1}^\top\hat{\boldsymbol{p}}' = -\mathbf{1}^\top\bar{\boldsymbol{p}}\,, \\ & \hat{\boldsymbol{p}}' \leq \hat{\boldsymbol{g}}^{avail} - \hat{\boldsymbol{d}}\,, \\ & \hat{\boldsymbol{p}}' \geq \min(-\hat{\boldsymbol{d}}, \hat{\boldsymbol{g}}^{avail})\,, \end{aligned} \tag{6.7}$$

where $\sum_{k\in\mathcal{I}_W}(p_k - p'_k)$ is the total curtailment. The linear optimisation depends only on the current demand profile and therefore is independent of time. In reality, a conventional generator requires time to accelerate/decelerate to its new power output. Therefore, the main assumption is that sufficient forecasting is performed in order to operate at the optimal solution. However, it may not be feasible to operate with maximum wind generation due to low system strength. Thus the linear optimisation may underestimate curtailment.

6.1.2 Minimise system change

The quadratic optimisation reschedules generation by minimising the change from the current set of power injections, $\boldsymbol{p}$. The main assumption that motivates this method is that the current power injections, $\boldsymbol{p}$, are the most economic. This is reasonable because in Section 6.2, the inputs are balanced by replacing expensive coal and gas generation with inexpensive wind generation, which replicates minimising operational costs. Therefore, the quadratic optimisation finds a new dispatch solution that is close to the most economic and satisfies transmission constraints. The quadratic objective function is given by

$$\min_{\hat{p}'_k} \sum_{k=1}^{n} \left[(\hat{p}'_k - \hat{p}_k)^2 - \lambda \mathbb{1}_{k\in\mathcal{I}_W}(\hat{p}'_k - \hat{p}_k)\right] , \tag{6.8}$$

where $\lambda > 0$ is the penalty applied to curtailing wind generation; $\lambda = 0$ corresponds to no penalty and results in more wind curtailment, whereas $\lambda \gg 0$ corresponds to a harsh penalty, similar to the linear optimisation problem. Whilst harsh penalties significantly reduce wind curtailment, the trade-off is that the change in other inputs may increase dramatically.

The constraints of the quadratic optimisation are the same as the linear optimisation (6.7). However, to solve the problem using software such as Matlab, we need to represent the objective function (6.8) in standard quadratic form. We apply a change of variables, $\boldsymbol{x} = \hat{\boldsymbol{p}}' - \hat{\boldsymbol{p}}$, in order to obtain a quadratic objective function given by

$$\min \ \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{x} - \lambda \cdot \mathbb{1}_{\mathcal{I}_W}^\top \boldsymbol{x} , \tag{6.9}$$

which is in standard quadratic form. The branch power flow constraints for the new variable, $\boldsymbol{x}$, are adapted from Equations (6.2) and (6.3) as follows:

$$\begin{aligned} -\mathbf{1} - \bar{H}\bar{\boldsymbol{p}}' - \hat{H}\hat{\boldsymbol{p}} \leq \ \hat{H}\hat{\boldsymbol{p}}' - \hat{H}\hat{\boldsymbol{p}} \ &\leq \mathbf{1} - \bar{H}\bar{\boldsymbol{p}}' - \hat{H}\hat{\boldsymbol{p}} , \\ \hat{H}\boldsymbol{x} \ &\leq \mathbf{1} - H\boldsymbol{p} , \qquad (6.10) \\ -\hat{H}\boldsymbol{x} \ &\leq \mathbf{1} + H\boldsymbol{p} . \qquad (6.11) \end{aligned}$$

The constraints (6.10) and (6.11) ensure that the absolute power flows from re-dispatch are within capacity, $|\boldsymbol{f}'| \leq \mathbf{1}$. In addition to the power flow constraint, we also derive the balancing equation that requires the sum of the power injections to be zero. Using the result from Equation (6.4), the balancing equation for the new variable, $\boldsymbol{x}$, is given by

$$\mathbf{1}^\top \boldsymbol{x} = \mathbf{1}^\top \hat{\boldsymbol{p}}' - \mathbf{1}^\top \hat{\boldsymbol{p}} = -\mathbf{1}^\top \bar{\boldsymbol{p}} - \mathbf{1}^\top \hat{\boldsymbol{p}} = -\mathbf{1}^\top \boldsymbol{p}\,. \tag{6.12}$$

Finally, the bounds from the linear optimisation (6.7) for the re-dispatch solution, $\hat{\boldsymbol{p}}'$, are translated by $-\hat{\boldsymbol{p}}$ in order to ensure the solution is feasible. The objective (6.9) and the constraints (6.10)-(6.12) form the following quadratic optimisation problem:

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \tfrac{1}{2}\boldsymbol{x}^\top \boldsymbol{x} - \lambda \cdot \mathbb{1}_{\mathcal{I}_w}^\top \boldsymbol{x} \\ \text{s.t.} \quad \hat{H}\boldsymbol{x} &\leq \mathbf{1} - H\boldsymbol{p}\,, \\ -\hat{H}\boldsymbol{x} &\leq \mathbf{1} + H\boldsymbol{p}\,, \\ \mathbf{1}^\top \boldsymbol{x} &= -\mathbf{1}^\top \boldsymbol{p}\,, \\ \boldsymbol{x} &\leq \hat{\boldsymbol{g}}^{avail} - \hat{\boldsymbol{d}} - \hat{\boldsymbol{p}}\,, \\ \boldsymbol{x} &\geq \min(-\hat{\boldsymbol{d}}, \hat{\boldsymbol{g}}^{avail}) - \hat{\boldsymbol{p}}\,. \end{aligned} \tag{6.13}$$

Suppose the optimisation has solution $\boldsymbol{x}^*$. Then, the re-dispatch solution is recovered from $\hat{\boldsymbol{p}}' = \boldsymbol{x}^* + \hat{\boldsymbol{p}}$.

6.2 Long-term congestion management model

Power transfers across the network are constrained by transmission capacity. New wind farms that are connected to renewable energy zones share the transmission capacity with the existing wind farms. As the transmission capacity is used, periods of very high wind output may be directed to curtail due to transmission constraints in the network.

Furthermore, the correlation of wind power exacerbates the frequency of congestion and curtailment, which increases short-term congestion management costs. Therefore, long-term planning is essential to ensure sufficient transmission capacity is available in order to avoid the most frequent congestions.

We develop a *long-term congestion management* model that quantifies the impact of additional wind capacity on network congestion and wind curtailment in SA. An overview of the model is described in Figure 6.1. Then, we briefly describe each step of the model, and refer to more detailed descriptions.

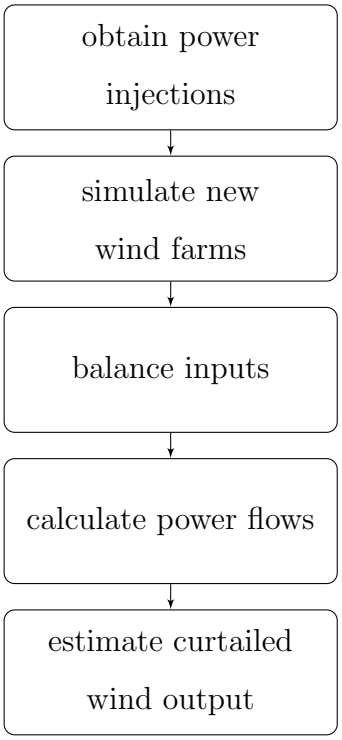


Figure 6.1: Flowchart of the long-term congestion management model.

Obtain power injections The power injections consist of the generation and load time series from the DC model. Initially, we balance the power injections for the AC model in Section 4.2, and then perform the DC approximation in Section 4.4. DC power flows assume a lossless network and therefore the sum of the power injections is zero.

Simulate new wind farms New hypothetical wind farms are incorporated into the existing network. We simulate the time series of generation for each new wind farm and combine this with the power injections. The simulation is performed using the wind power simulation explained in Chapter 5.

Balance inputs Incorporating the additional simulated wind generation into the power injections, increases total generation whilst total demand remains constant. This may create an unrealistic mismatch between total generation and total demand. Therefore, we offset the additional wind generation by removing existing coal and gas generation. This process replicates the evolution of the power network from conventional power to renewable power and is assumed to be the most economic. First, coal generation is removed because SA's Northern Power Station closed in 2016. Second, gas generation is removed because gas is the most expensive energy resource. However, the gas turbines in SA are important for system strength.

Therefore, we maintain a minimum of 150 MW of gas power.

Calculate power flows Power flow calculations are performed in MATPOWER. This is a trusted software used by power system researchers from academia, government, and industry [67, 68]. We use a closed-form DC power flow. Whilst the DC power flow is not as accurate as AC calculations for specific time points, it has proven to be a useful analysis tool in identifying overall trends [12, 35, 63].

Estimate curtailed wind output The most economic power injections may produce power flows that exceed transmission constraints. When this occurs, generation is rescheduled in order to obtain a re-dispatch solution that satisfies the transmission constraints. Then, the total amount of wind curtailment is estimated from the re-dispatch solution. We use the linear optimisation from Section 6.1.1 because we believe that operators will be able to implement sufficient forecasting and other clever solutions at their disposal in order to operate at or near maximum wind generation.

6.3 Results

Wind farms in SA are predominantly located in two renewable energy zones: Mid-North and South East. As described in Chapter 5, we separate the Mid-North into an East and West component in order to increase the accuracy of the wind power simulation. Then, we perform a load-flow analysis of the SA power network where additional wind is incorporated into one of the three zones: South East, Mid-North (East) and Mid-North (West). To clarify, one simulation incorporates 100 MW of additional wind capacity in the South East. A list of the simulations are displayed below.

Zone	Wind Capacity	Connection
South East	100 - 600 MW	Snuggery
Mid-North (East)	100 - 600 MW	Belalie
Mid-North (West)	100 - 600 MW	Brinkworth 132

Table 6.1: Simulations of additional wind capacity at the corresponding connection point in SA. A total of 18 simulations are conducted at 100 MW intervals.

Each simulation consists of the time points $t = 1, \ldots, 105120$, corresponding to

every 5-minutes of the year 2015. For each time point, we record the pre-dispatch, $\boldsymbol{p}_t$, corresponding to the power injections after Step 3 of the flowchart in Figure 6.1, the resulting power flow of the pre-dispatch, $\boldsymbol{f}_t$, and the re-dispatch, $\boldsymbol{p}'_t$, from Step 5. The total amount of curtailed wind output for the year is given by

$$\sum_t \sum_{k \in \mathcal{I}_W} (g_{k,t} - g'_{k,t}) \,,$$

where $g_{k,t}$ and $g'_{k,t}$ is the generation at bus k at time t in the pre-dispatch and re-dispatch, respectively. Additionally, the wind penetration for the year is given by

$$\frac{\text{total generation from wind}}{\text{total generation}} = \frac{\sum_t \sum_{k \in \mathcal{I}_W} g'_{k,t}}{\sum_t \sum_k g'_{k,t}} \,.$$

In the subsequent sections, we use our results to provide insight into future network congestion and wind curtailment in SA.

6.3.1 Optimal renewable energy zone

For our purpose, we are comparing only the South East and Mid-North. The ISP [47] outlines a number of factors that determine the optimal REZ for wind farm integration. However, we take a simpler approach by comparing wind penetration curves given the wind capacity investment in each zone. Wind penetration is a reasonable measure as it depends on available transmission and quality of the wind resource, both of which are considered in the ISP. From our simulations, we construct the wind penetration curves displayed in Figure 6.2, and provide the numerical values in Table 6.2.

It is evident, that for the same amount of wind capacity, the Mid-North has much higher penetration in comparison to the South East. Therefore, the optimal renewable energy zone is the Mid-North. Moreover, the Mid-North (West) appears to have the best quality of wind resource because it has the largest amount of available wind generation for the same amount of capacity. However, at higher additional wind capacities, wind generation in the Mid-North (West) is curtailed, resulting in constrained penetration.

6.3.2 Estimate spare network capacity

AEMO defines spare network capacity by the MW value of additional generation that can be transported from the REZ to the required load centre [47]. This is

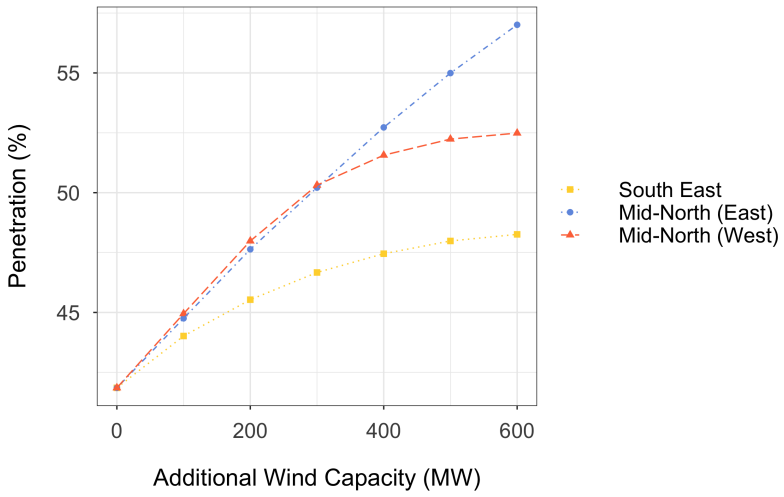


Figure 6.2: The percentage of wind power that penetrates the network given the additional investment in wind capacity. Investment in wind capacity in the Mid-North results in higher wind penetration levels.

an important limit because if the total wind capacity in a REZ exceeds the spare network capacity then curtailment will occur. Therefore, we estimate the spare network capacity by incorporating additional wind capacity until curtailment occurs. We plot the resulting curtailment against the amount of additional wind capacity in Figure 6.3, and provide the numerical values in Table 6.2.

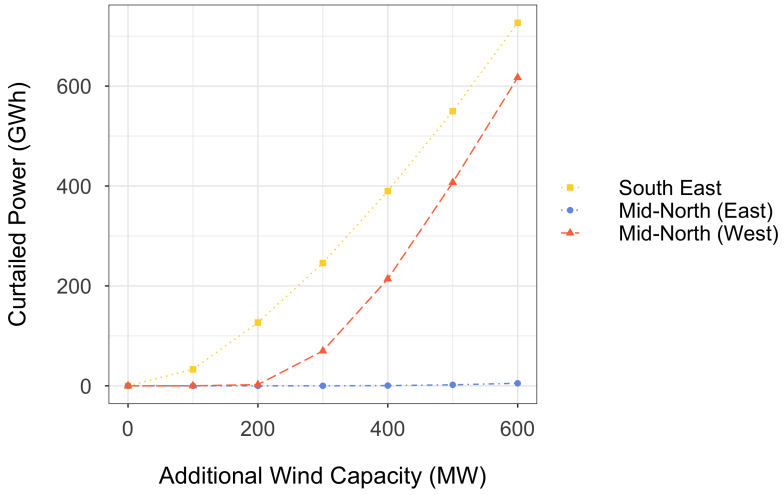


Figure 6.3: The total amount of curtailed wind power (GWh) given the additional investment in wind capacity. Immediate curtailment is evident in the South East.

Mid-North It is immediately evident that the Mid-North has significantly more capacity than the South East. Furthermore, curtailment in the Mid-North (East)

and Mid-North (West) begins at 300 MW and 200 MW, respectively. Therefore, we estimate the spare network capacity to be approximately 500 MW. Having said that, it is reasonable for operators to tolerate a small amount of curtailment in order to increase penetration. Since wind curtailment in the Mid-North (East) is negligible, the network can tolerate 900 MW of additional wind capacity and experience less than 3% curtailment of the additional generation.

South East Immediate curtailment is evident for the South East, and therefore the spare network capacity is less than 100 MW. This differs dramatically from the 2018 ISP which reports the spare network capacity to be 500 MW. We note that the spare network capacity of the REZs is used as an assumption to the ISP modelling.

We identify two changes in the physical network: our simplified network combines the Mayura and Snuggery substations, and the Heywood inter-connector received a substantial transmission upgrade in 2016. The Mayura and Snuggery substations are geographically adjacent and would have only a minimal effect on the power flows. Additionally, in Section 6.3.3, we show that other network elements impose the constraint prior to the inter-connector and therefore it is not the reason for increased spare network capacity in the ISP.

Furthermore, in our case study additional wind capacity is connected to Snuggery, which already connects 335 MW of wind capacity from the Canunda and Lake Bonney wind farms. We chose the Snuggery substation due to the increased wind resource near the coast, which capitalises on the prevailing south-west coastal winds, and therefore new wind farms are likely to be built closer to the coast. Whilst changing the connection to Blanche or Mt Gambier may reduce curtailment, our modelling suggests this will not achieve a spare network capacity of 500 MW. We expect that AEMO has access to more accurate data, and therefore this may indicate that our physical model could be improved.

	Additional Wind Capacity (MW)						
	0	100	200	300	400	500	600
	Additional available wind generation (GWh)						
South East	-	252.94	505.22	754.92	1006.55	1261.09	1516.29
Mid-North (East)	-	294.64	593.44	877.49	1173.89	1468.08	1760.75
Mid-North (West)	-	314.04	629.79	954.64	1267.69	1592.03	1904.21
	Additional wind generation after curtailment (GWh)						
South East	-	219.93	378.52	509.10	616.73	711.38	789.55
Mid-North (East)	-	294.64	593.44	877.44	1173.40	1466.04	1755.60
Mid-North (West)	-	314.04	627.29	884.80	1053.64	1184.99	1287.06
	Percentage of additional wind generation curtailed (%)						
South East	-	13.05	25.08	32.56	38.73	43.59	47.93
Mid-North (East)	-	0.00	0.00	0.00	0.04	0.14	0.29
Mid-North (West)	-	0.00	0.40	7.32	16.89	25.57	32.41
	Wind penetration (%)						
South East	41.85	44.02	45.53	46.67	47.45	47.98	48.26
Mid-North (East)	41.85	44.75	47.64	50.21	52.73	54.99	57.01
Mid-North (West)	41.85	44.95	47.98	50.31	51.57	52.24	52.49
	Curtailed wind output (GWh)						
South East	0.00	33.01	126.69	245.81	389.82	549.71	726.74
Mid-North (Eest)	0.00	0.00	0.00	0.04	0.49	2.04	5.14
Mid-North (West)	0.00	0.00	2.50	69.84	214.05	407.04	617.15

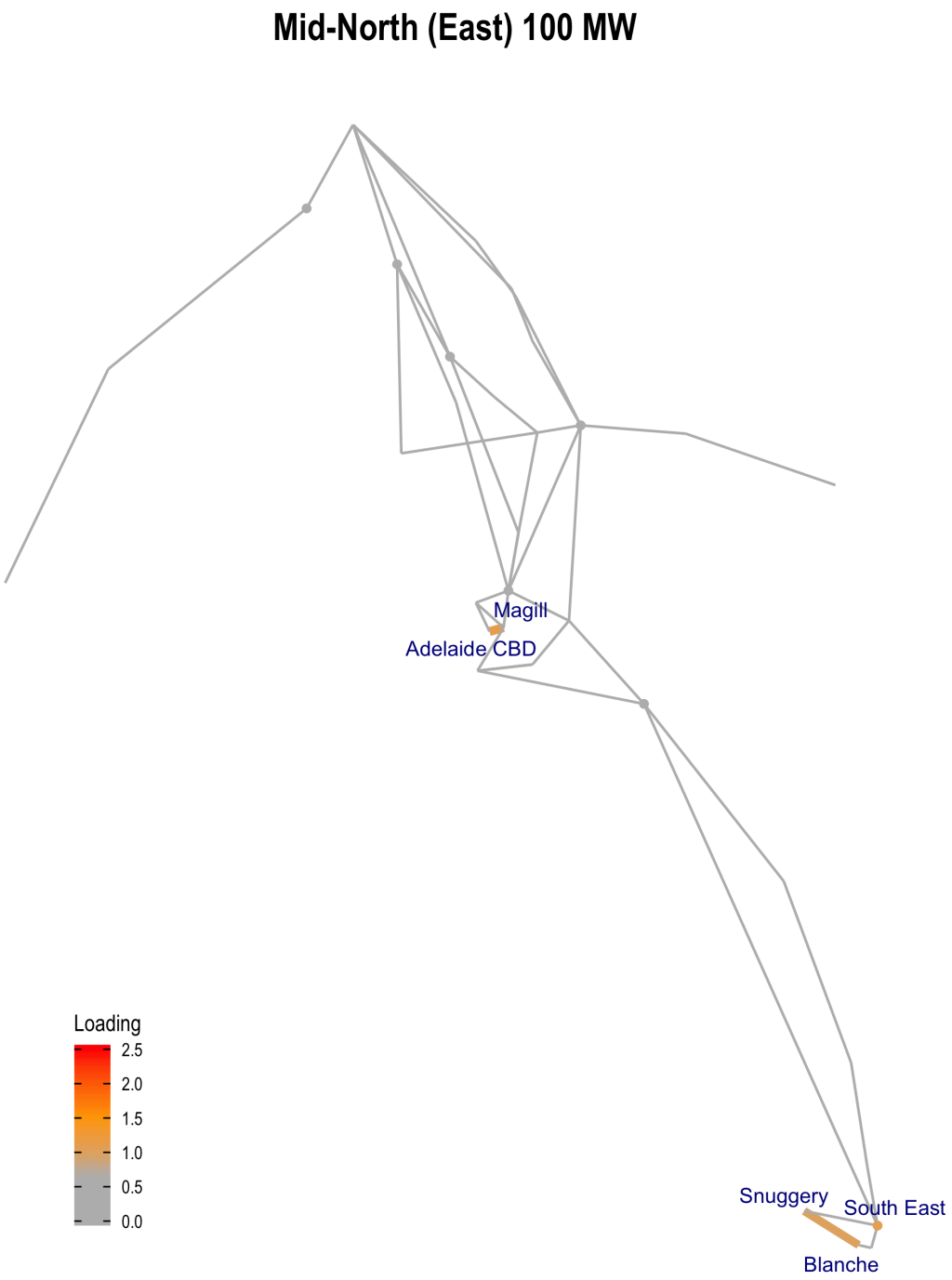
Table 6.2: Numerical values of additional wind generation for each simulation. This includes the values of wind penetration and curtailment from Figures 6.2 and 6.3.

6.3.3 Identify transmission constraints

Incorporating additional generation in a particular area of the network will cause a non-uniform increase in the power flows. Moreover, the change in power flow is governed by the PTDFs. This may lead to power flows that exceed transmission capacity in the pre-dispatch solution and need to be rescheduled, potentially causing curtailment. In order to identify the transmission elements that constrain wind penetration, we investigate the maximal loading of the power flows, $|\boldsymbol{f}_t|$, in Figures 6.4-6.6. Additionally, we provide the numerical values of the maximal loadings in Table 6.3.

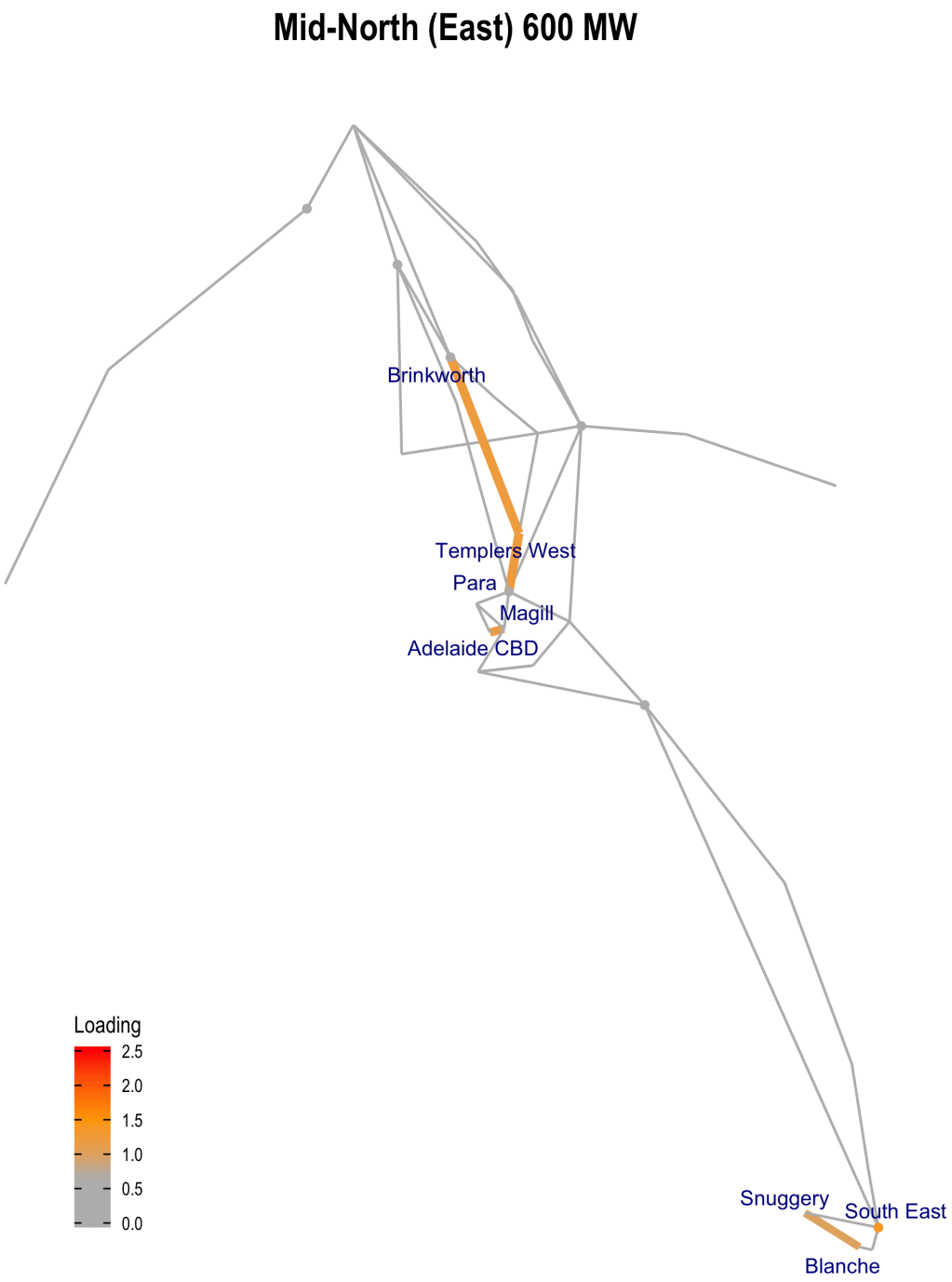
Mid-North In Figure 6.4a, the highlighted congested elements in the pre-dispatch do not result in curtailment because curtailment occurs beyond 200 MW as shown in Figure 6.3. These correspond to infrequent congestions that are rectified by generation re-dispatch. However, as additional wind capacity is increased in the Mid-North (East), Figure 6.4b demonstrates that the congestions in the transmission lines *Brinkworth - Templers West* and *Templers West - Para 275* emerge. This is also a feature of the Mid-North (West) in Figure 6.5b, with the addition of the *Brinkworth transformer*. In Section 6.3.4, we assess the benefit of connecting to the Bungama bus instead of Brinkworth.

South East Once again, immediate congestion is present in the South East. From Section 6.3.2, we know this congestion induces wind curtailment. Congested elements include the 132 kV transmission lines *Snuggery - South East 132*, *Snuggery - Blanche*, *Blanche - Mt Gambier* and the *South East transformer*. Therefore, congestion occurs before transmission across the inter-connector, eliminating this to be a possible reason that explains the discrepancy between our estimated spare network capacity for the simplified network in 2015 and AEMO's reported spare network capacity in the 2018 ISP. Furthermore, as additional wind capacity is increased to 600 MW, numerous transmission elements become congested, thus making it overly expensive to upgrade transmission to increase the spare network capacity to values much larger than 100 MW. We explore this further in Section 6.3.5.



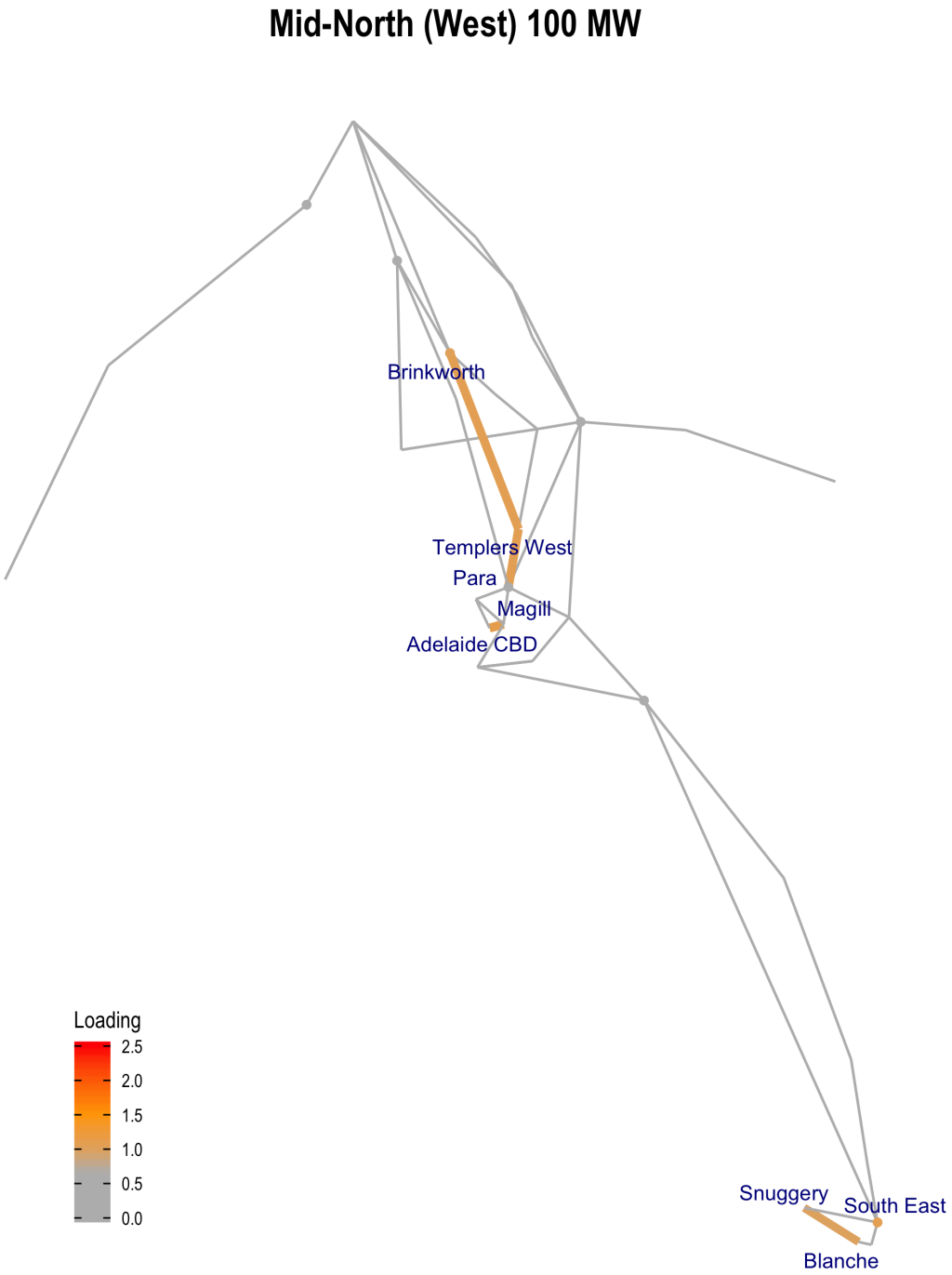
(a) Incorporate 100 MW of additional wind capacity into the Mid-North (East).

Figure 6.4: Maximum loading of the transmission network in the pre-dispatch solution when connecting additional wind capacity in the Mid-North (East). The maximal loading of the network elements is less than 1.2 pu, and does not result in wind curtailment.



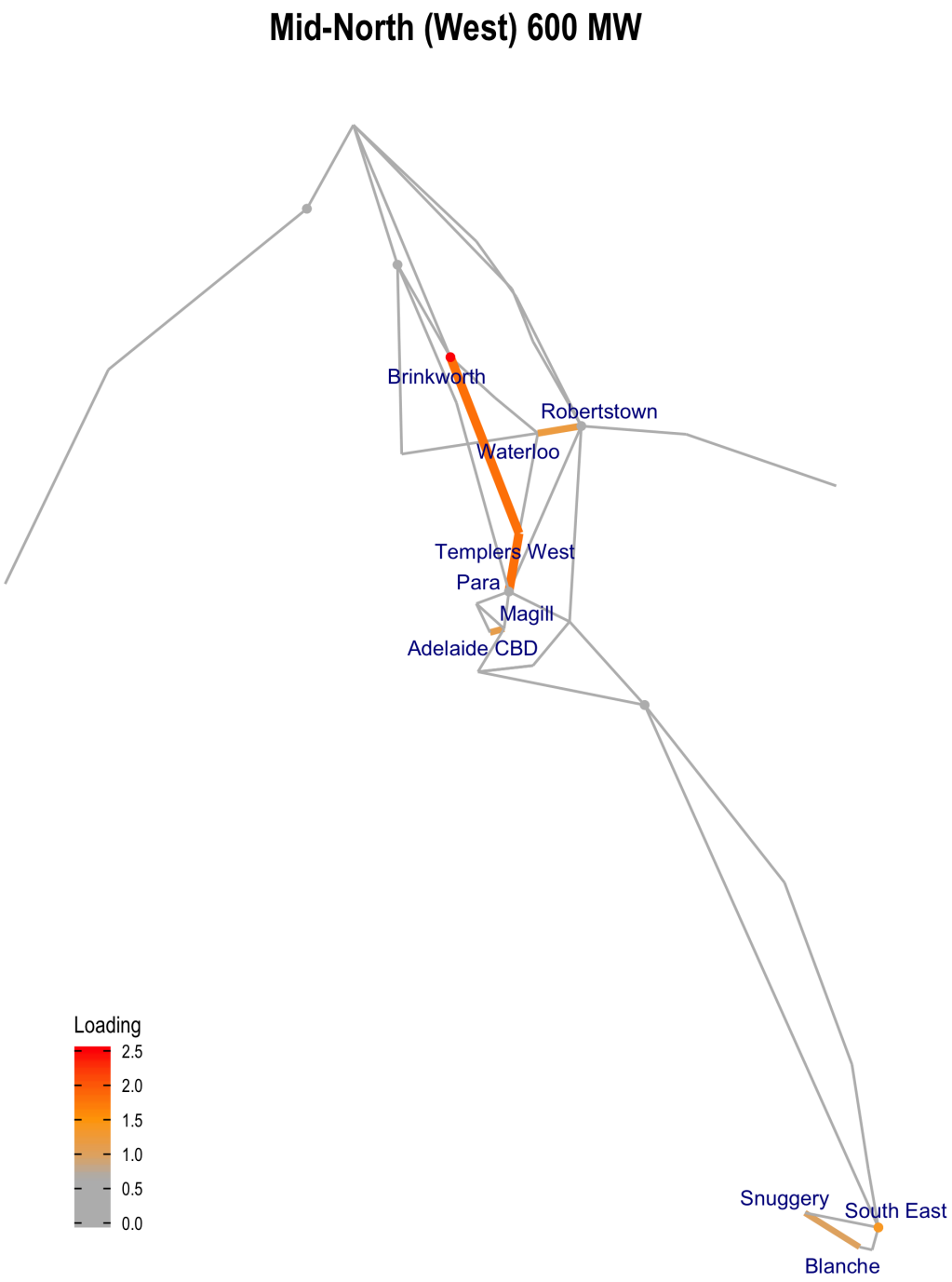
(b) (cont.) Incorporate 600 MW of additional wind capacity into the Mid-North (East).

Figure 6.4 (cont.): Maximum loading of the transmission network in the pre-dispatch solution when connecting additional wind capacity in the Mid-North (East). The main congestions have a maximal loading of 1.22 pu, and occur in series between Brinkworth 275 and Para 275.



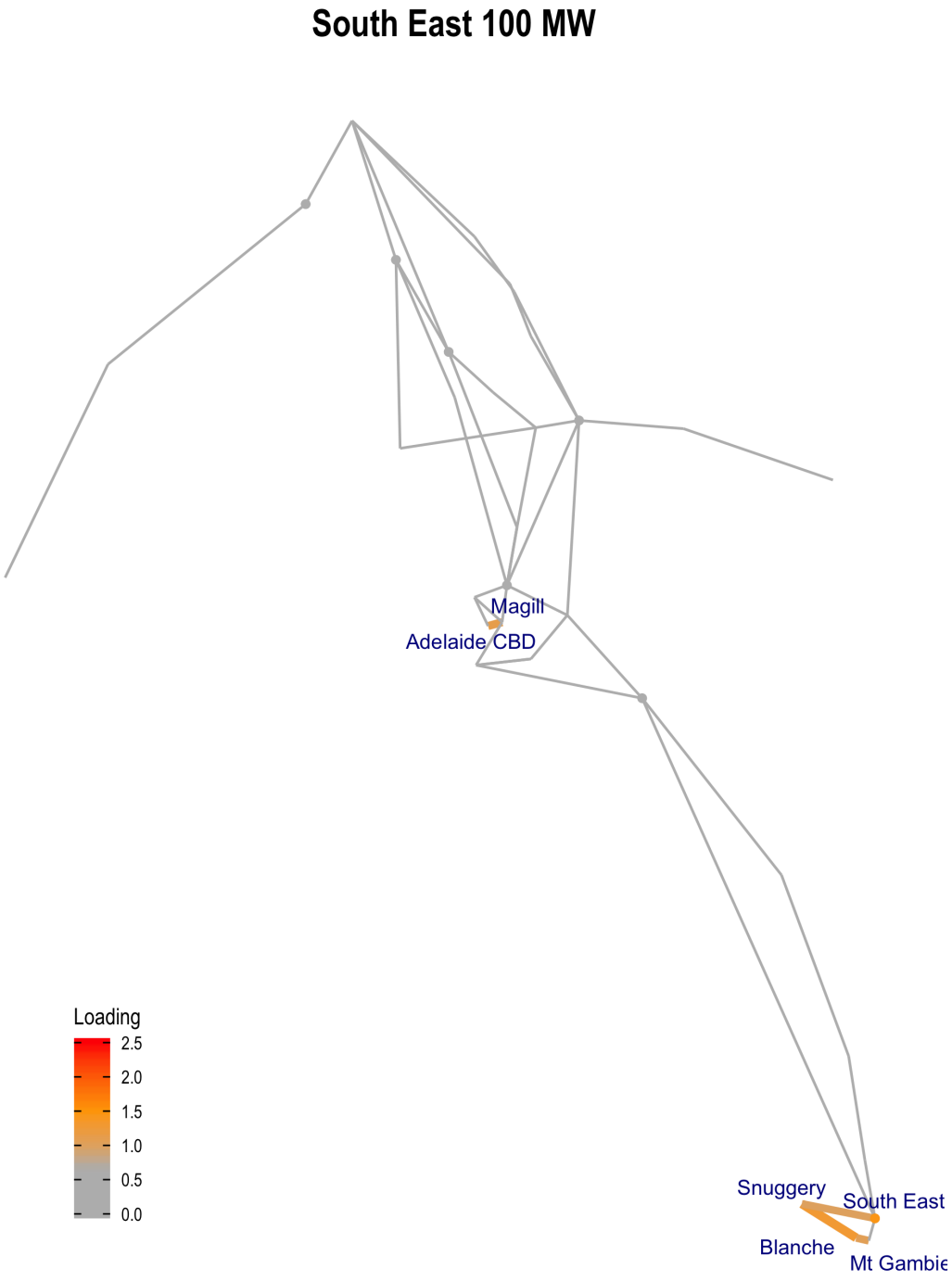
(a) Incorporate 100 MW of additional wind capacity into the Mid-North (West).

Figure 6.5: Maximum loading of the transmission network in the pre-dispatch solution when connecting additional wind capacity in the Mid-North (West). The main congestions between Brinkworth 132 and Para 275 have a maximal loading of less than 1.1 pu, and do not result in wind curtailment.



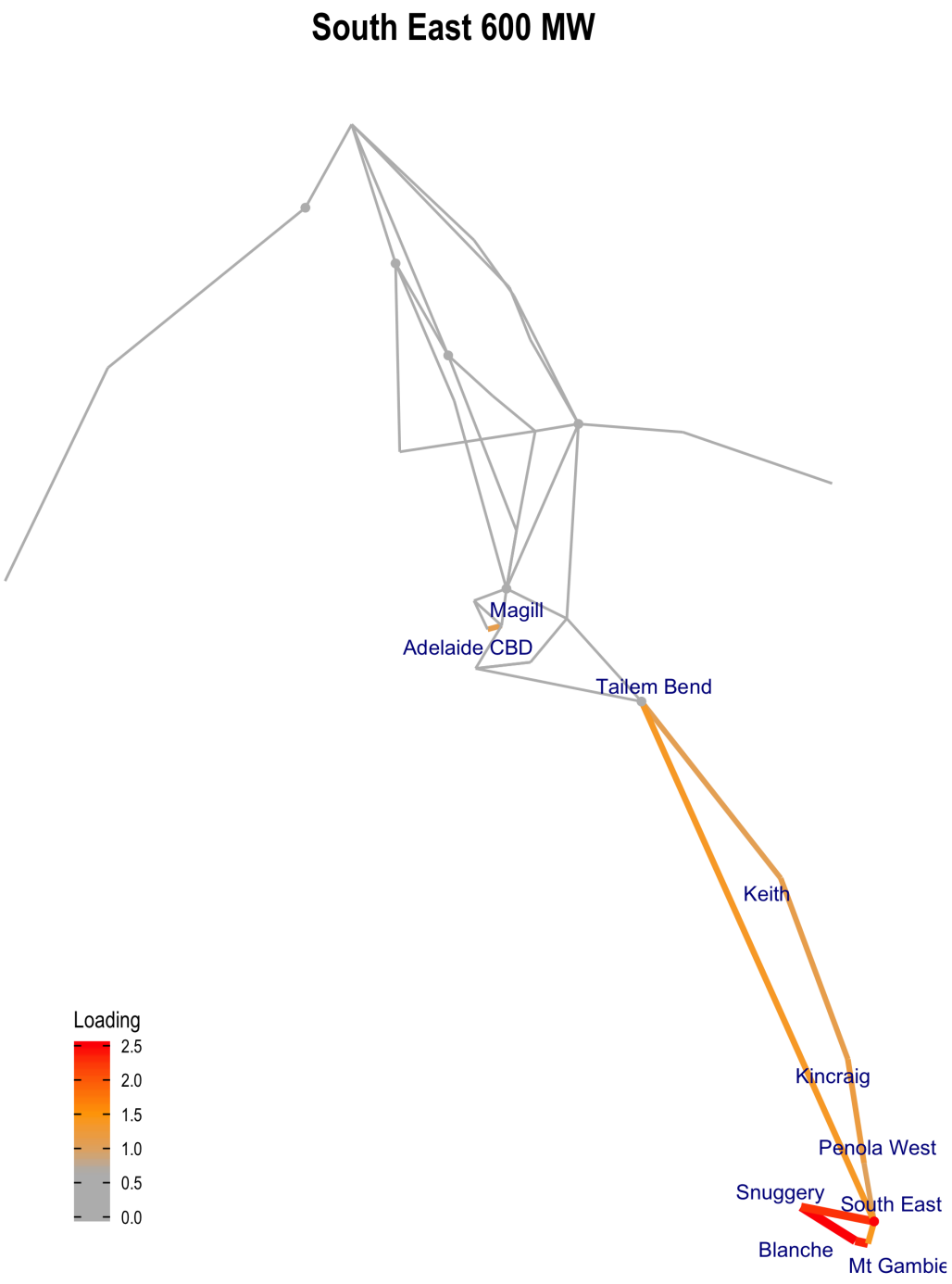
(b) (cont.) Incorporate 600 MW of additional wind capacity into the Mid-North (West).

Figure 6.5 (cont.): Maximum loading of the transmission network in the pre-dispatch solution when connecting additional wind capacity in the Mid-North (West). The main congestions between Brinkworth 132 and Para 275 are similar to the Mid-North (East) except more severe with loadings above 1.8 pu.



(a) Incorporate 100 MW of additional wind capacity into the South East.

Figure 6.6: Maximum loading of the transmission network in the pre-dispatch solution when connecting additional wind capacity in the South East. The maximal loading of the network elements is between 1-1.5 pu, and results in significant wind curtailment.



(b) (cont.) Incorporate 600 MW of additional wind capacity into the South East.

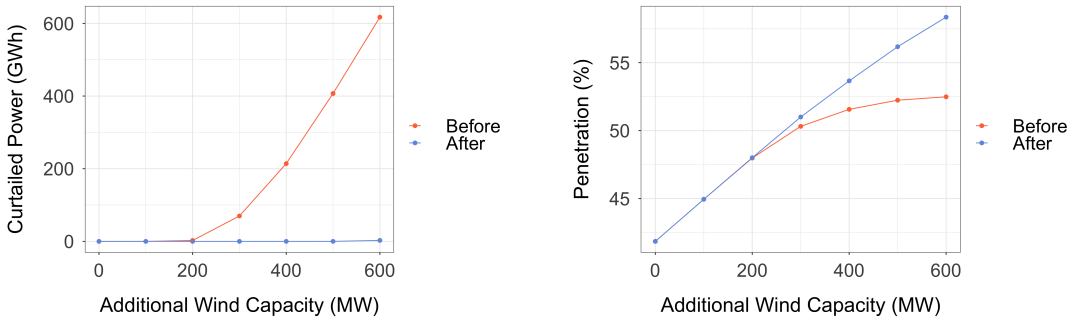
Figure 6.6 (cont.): Maximum loading of the transmission network in the pre-dispatch solution when connecting additional wind capacity in the South East. A significant amount of additional wind capacity in the South East results in many congested network elements.

	Branch		Additional Wind Capacity (MW)					
	to	from	100	200	300	400	500	600
Mid-North (East)	Snuggery	Blanche	1.00	1.00	1.00	1.00	1.00	1.00
	Tailem Bend	Keith	0.52	0.52	0.52	0.52	0.51	0.51
	Keith	Kincraig	0.57	0.57	0.57	0.57	0.56	0.56
	Mt Gambier	Blanche	0.80	0.80	0.80	0.80	0.80	0.80
	Kincraig	Penola West	0.64	0.64	0.64	0.64	0.63	0.63
	Penola West	South East 132	0.50	0.50	0.50	0.50	0.50	0.50
	South East 132	Mt Gambier	0.42	0.42	0.42	0.42	0.42	0.42
	Snuggery	South East 132	0.75	0.75	0.75	0.75	0.75	0.75
	Waterloo	Robertstown 132	0.65	0.63	0.62	0.62	0.59	0.62
	Magill	Adelaide CBD	1.13	1.13	1.13	1.13	1.13	1.13
	Templers West	Para 275	0.99	0.98	0.98	1.00	1.11	1.22
	Tailem Bend 275	South East 275	0.76	0.76	0.76	0.76	0.84	0.94
	Brinkworth 275	Templers West	0.99	0.98	0.98	1.00	1.11	1.22
	Brinkworth 132	Brinkworth 275	0.77	0.77	0.76	0.77	0.80	0.83
	South East 132	South East 275	1.08	1.13	1.19	1.25	1.30	1.39
Mid-North (West)	Snuggery	Blanche	1.00	1.00	1.00	1.00	1.00	1.00
	Tailem Bend	Keith	0.52	0.52	0.52	0.52	0.51	0.52
	Keith	Kincraig	0.57	0.57	0.57	0.56	0.56	0.56
	Mt Gambier	Blanche	0.80	0.80	0.80	0.80	0.80	0.80
	Kincraig	Penola West	0.64	0.64	0.64	0.63	0.63	0.63
	Penola West	South East 132	0.50	0.50	0.50	0.50	0.50	0.49
	South East 132	Mt Gambier	0.42	0.42	0.42	0.42	0.42	0.42
	Snuggery	South East 132	0.75	0.75	0.75	0.75	0.75	0.75
	Waterloo	Robertstown 132	0.76	0.85	0.93	1.02	1.12	1.20
	Magill	Adelaide CBD	1.13	1.13	1.13	1.13	1.13	1.13
	Templers West	Para 275	1.07	1.15	1.23	1.43	1.69	1.81
	Tailem Bend 275	South East 275	0.76	0.76	0.76	0.76	0.82	0.99
	Brinkworth 275	Templers West	1.07	1.15	1.23	1.43	1.69	1.81
	Brinkworth 132	Brinkworth 275	1.03	1.31	1.57	1.86	2.14	2.47
	South East 132	South East 275	1.08	1.13	1.18	1.26	1.32	1.38
South East	Snuggery	Blanche	1.29	1.58	1.87	2.16	2.45	2.74
	Tailem Bend	Keith	0.62	0.73	0.83	0.92	1.00	1.08
	Keith	Kincraig	0.67	0.77	0.87	0.96	1.05	1.13
	Mt Gambier	Blanche	1.06	1.31	1.57	1.83	2.08	2.34
	Kincraig	Penola West	0.74	0.84	0.94	1.03	1.12	1.20
	Penola West	South East 132	0.59	0.69	0.79	0.87	0.95	1.02
	South East 132	Mt Gambier	0.57	0.74	0.91	1.08	1.25	1.42
	Snuggery	South East 132	1.00	1.25	1.51	1.76	2.01	2.26
	Waterloo	Robertstown 132	0.68	0.68	0.68	0.69	0.70	0.70
	Magill	Adelaide CBD	1.13	1.12	1.12	1.12	1.11	1.12
	Templers West	Para 275	0.99	0.99	0.99	0.99	0.99	0.99
	Tailem Bend 275	South East 275	0.88	1.01	1.15	1.24	1.29	1.38
	Brinkworth 275	Templers West	0.99	0.99	0.99	0.99	0.99	0.99
	Brinkworth 132	Brinkworth 275	0.74	0.72	0.70	0.70	0.69	0.69
	South East 132	South East 275	1.46	1.92	2.36	2.82	3.29	3.73

Table 6.3: Maximal loading of the transmission network in the pre-dispatch solution when connecting additional wind capacity. Branches that do not exceed capacity are omitted.

6.3.4 Investigate the connection to the Mid-North

We found that connecting the additional wind generation to Brinkworth in the Mid-North (West) resulted in congestions in the transmission lines *Brinkworth - Templers West* and *Templers West - Para 275* as well as the *Brinkworth transformer*. Therefore, we observe the benefit of connecting additional wind generation to Bungama. We choose Bungama because there are three parallel paths connecting Bungama to Adelaide and therefore the power flows are less sensitive to the power injection. This can be seen in the visualisation of the power transmission distribution factors from Section 4.5. We compare the resulting curtailment and penetration in Figure 6.7.



(a) Curtailed wind output (GWh). (b) Wind penetration.

Figure 6.7: Curtailed wind power output (left) and wind penetration (right) from connecting to Brinkworth (before) and Bungama (after) in the Mid-North (West). The connection to Bungama eliminates the previous curtailment observed from the connection to Brinkworth.

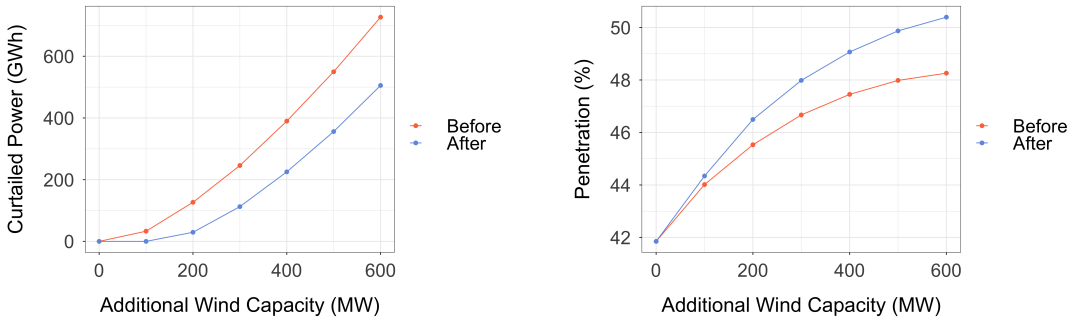
It is evident that particular connections have larger tolerances to large power injections than others. In this case, Bungama is the superior connection for wind farm integration as it eliminates all previous curtailment, seen in Figure 6.7a. Using this connection point, the estimated spare network capacity in the Mid-North would increase to approximately 1000 MW with negligible curtailment. This highlights that the spare network capacity is highly dependent on the point of connection, and motivates modelling techniques like ours where we have control over the connection point.

6.3.5 Alleviate curtailment in the South East

In Section 6.3.3, we identify four transmission branches that constrain spare network capacity in the South East. As suggested by our model, we perform the following transmission upgrades in order to eliminate curtailment within the first 100 MW of additional wind capacity:

- increase the capacity of *South East transformer* by 50%.
- increase the capacity of *Snuggery - Blanche* transmission line by 30%.
- increase the capacity of *Blanche - Mt Gambier* transmission line by 5%.
- increase the capacity of *Snuggery - South East* transmission line by 0.5%.

We compare the results of before and after transmission upgrades in Figure 6.8.



(a) Curtailed wind output (GWh). (b) Wind penetration.

Figure 6.8: Curtailed wind power output (left) and wind penetration (right) of before and after transmission upgrades in the South East. The proposed upgrades eliminate curtailment within the first 100 MW of additional wind capacity.

It is evident in Figure 6.8a that the proposed transmission upgrades eliminate curtailment for the first 100 MW of additional wind capacity. However, the spare network capacity is only increased to a value between 100-200 MW, corresponding to less than 1% increase in wind penetration. After which, curtailment occurs at a relatively steep rate.

6.4 Discussion

In this chapter we brought together models from previous chapters to form our long-term congestion management model for the SA power network. The exact,

real-life network is not available and therefore we developed a simplified open-access AC model from publicly available data. There were two main challenges with this model. First, public demand data is only available for the distribution network, which omits power supplied directly to industry and confidential loads. This results in an unrealistic mismatch between generation and load which requires a lot of work to balance. Second, there is very limited published data about FACTS devices and therefore expert knowledge is required to handle reactive power supplies. Furthermore, we used a DC approximation in order to obtain a closed-form power flow solution. We briefly assessed the usefulness of the DC approximation to our AC model. However, our assessment relies on the results of Purchala *et al.* [54], and is by no means exhaustive. We recognise a definite need for future work into the developing and testing of network models. This will provide consistency and allow other research problems to avoid the time consuming process of network estimation.

We limited ourselves to using the time series of generation and load from 2015 in order to avoid wind curtailment that occurred after September 2016 when SA experienced a state-wide blackout. If curtailment were present in the original data this would introduce a whole new set of difficulties: curtailed wind farms would affect the estimation of regression parameters in the wind power simulation and the upper bound in the re-dispatch model would need to be able to distinguish between curtailed and normal operation. The ability to handle curtailed data is crucial for the study of more recent years and hence, handling curtailed data is an important area of future research.

Our modelling is restricted to investigating additional wind capacity in the Mid-North and South East. Whilst this may seem fairly restrictive, these are the two major REZs that cover the majority of the network. Although as the spare network capacity within the Mid-North is used, simulation of wind generation in other areas will become of interest. Potentially, more flexibility could be achieved through spatio-temporal modelling which is another area of future research.

The estimation of wind curtailment relies on the short-term re-dispatch model. We used a novel short-term re-dispatch model in the form of a linear optimisation which maximises wind generation. The re-dispatch may underestimate curtailment for three reasons. First, the optimisation is independent of time and therefore assumes sufficient forecasting is performed to be able to operate at the optimal solution. Second, in some circumstances, operating at maximum wind generation is not cost effective due to conventional generator shut down and start up costs. Third, the only security constraint applied corresponds to the transmission constraints. Others

include $N-1$ security which ensures the network remains stable after the loss of any one element. Adding these additional constraints will increase the estimated amount of wind curtailment. Similar to the network model, the real re-dispatch is not available and therefore there is a need for future work into developing and testing re-dispatch models.

We have developed a high-level model that provides critical information for specific wind farm integration problems. In Section 6.3.4, we saw that choosing a different bus connection for additional wind generation in the Mid-North (West) dramatically changed network congestion. There are fewer connection options for other zones such as the South East. However, different connections may reduce the number of required line upgrades for curtailment free integration. This motivates a need for an efficient exhaustive model that determines an optimal connection arrangement, but for now this is yet another area for future work.

Chapter 7

Conclusion

In this thesis we have made two main contributions: an improved emergent cascade model and a new long-term congestion management model. The congestion management model consists of three new components, including a network model of SA, a wind power simulation and a short-term generation re-dispatch model. Here we recap the main findings of each section, discuss lessons learned, and describe future work.

Chapter 3: Emergent cascades

Findings In this chapter we discuss the emergent cascade model of Nesti *et al.* [38], which is the initial model that investigates cascades that arise from fluctuations in renewable energy. The authors use hourly power generation data of wind and solar farms to estimate the steady-state covariance matrix of the power injections. We find that the wind model has a reasonable goodness-of-fit and show that the diagnostics of the solar model suggest the presence of significant cycles. Using a power spectrum, we identify significant cycles in solar generation corresponding to the annual and daily cycles. We find that the extended solar model that incorporates a Fourier series fitted to the significant cycles improves the goodness-of-fit. Furthermore, we find that our extended model reduces the diagonal elements of the line covariance matrix and hence decreases the probability of line failure.

Lessons learned and future work The failure probabilities rely heavily on the estimation of the steady-state covariance matrix of the power injections. More accurate models for the steady-state covariance matrix, such as those developed in

this thesis, should be used. Furthermore, future work would be to estimate the steady-state covariance matrix of the power injections from shorter-time-interval data, such as 5-15 minutes, for more accurate estimates of the probabilities of line failure.

Chapter 4: South Australian power network

Future work In this chapter we outline our simplified open-access AC model of the South Australian transmission network and show its capability of emulating network operation. We then apply a DC approximation to obtain a closed-form power flow solution. Unfortunately, the public data used for our SA network model may not be realistic, and therefore applying our method to more accurate data, potentially obtained from the network operators ElectraNet, is future work. We note that a consequence of more accurate data may result in the network model becoming confidential as opposed to open-access.

Another area of future research is extending our method to incorporate AC power flows. This is inherently more difficult due to reactive power and requires a compatible AC re-dispatch model.

Chapter 5: Wind power simulation

Findings In this chapter we develop a wind power simulation for future wind farms in SA. Our model takes account of key features such as correlation with existing wind farms, climatic inputs, temporal dependence in wind generation and non-Gaussian errors. The model has been verified on existing wind generation data and shown to be accurate.

Lessons learned and future work Currently, the wind power simulation model is restricted to the South East and the Mid-North. Spatio-temporal modelling may be applied to the existing data in order to simulate the time series of generation for a new wind farm in other REZs. This becomes an important area of future work as the spare network capacity in the Mid-North is used.

Chapter 6: Congestion and wind curtailment

Findings Our last main section focuses on the long-term congestion management model. Integrating the wind farm simulation into a network model with a re-dispatch model provides a realistic method for modelling new wind farms at a high level of resolution. Since our open-access network model may not be realistic, our analysis is a proof of concept which demonstrates that comprehensive realistic models can describe future congestions. This information may aid network planning in long-term congestion management.

Lessons learned and future work We use a novel re-dispatch model designed to maximise wind generation. This may underestimate curtailment for two reasons. First, the re-dispatch is time independent and assumes sufficient forecasting is performed to operate at the optimal solution. However, maximum wind generation may not always be cost effective due to conventional generator start up and shut down costs. Second, we can further constrain the re-dispatch by including generator ramping rates and $N-1$ security. More sophisticated re-dispatch models will increase estimated wind curtailment and more accurately represent real operation. Introducing more sophisticated re-dispatch models is more future work.

Another area of future research is to be able to use generation data where curtailment is present. We were restricted to 2015 in order to avoid curtailment, but modelling of more recent years will need to manage curtailment. What makes this a very difficult problem is that curtailment is not published. Therefore, identifying wind farms that are curtailed is an important area of future research.

Closing remarks

Here we have provided an in-depth long-term congestion management model that quantifies the impact of connecting new wind farms on network congestion and wind curtailment in South Australia. We hope our methods will be used for future work investigating network congestion and wind curtailment in South Australia and elsewhere.

Appendix A

Revised South Australian network topology

We simplify the network topology of South Australia by combining various substations. We sum the load and generation at the individual buses to create the time series of load and generation for our simplified network. In Figure A.1, we provide a visualisation of the combined substations. In Table A.1, we present the edge-list and data for our simplified network.

In Chapter 4, we discussed that to emulate real operation, the generators are connected to the high-voltage buses via transformers. These transformers are necessary to control the voltage fluctuations of the high-voltage buses. In Figure A.2, we present a visual representation of the edge-vertex incidence matrix for the SA network with the additional transformers.

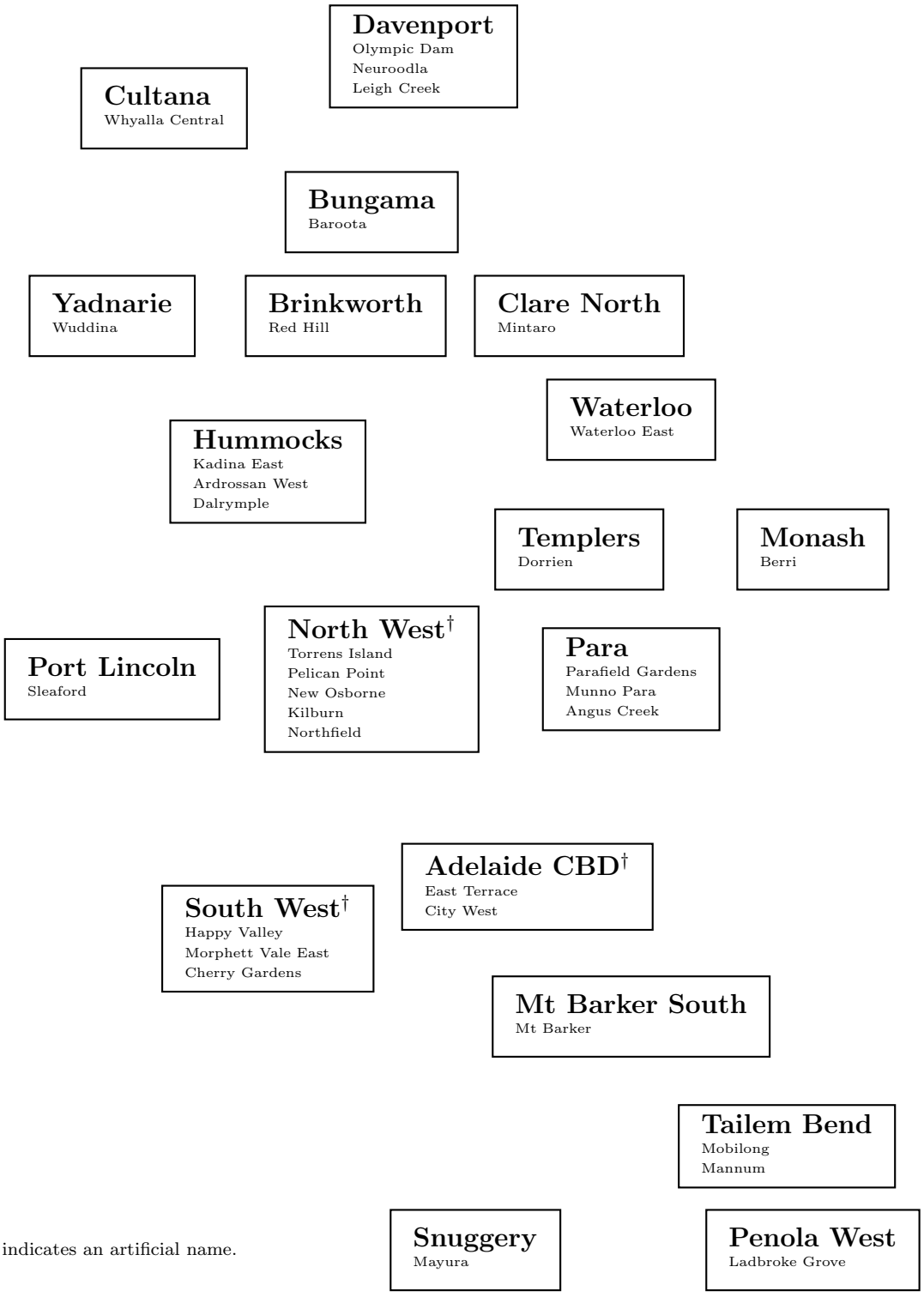


Figure A.1: Buses that are combined in order to obtain the simplified topology in Figure 4.2.

to	from	ID	voltage (kV)	Spr/Aut	Summer	Winter	Length (km)
Blanche	Snuggery	SNGBLA_1	132	111.00	105.00	126.00	49.60
Brinkworth 132	Bungama 132	REDBUN_T	132	150.00	141.00	173.00	40.83
Brinkworth 132	Clare North	BRICLN_1	132	160.00	160.00	160.00	31.88
Clare North	Waterloo	MINWAT_1	132	189.00	183.00	207.00	18.88
Cultana 132	Yadnarie	CULMBK_T	132	89.00	73.00	126.00	153.44
Hummocks	Bungama 132	BUNHUM_T	132	111.00	105.00	126.00	100.91
Keith	Tailem Bend 132	TABKEI_2	132	183.00	178.00	183.00	133.24
Kincraig	Keith	KEIKIN_1	132	183.00	183.00	183.00	109.39
Monash	North West Bend	NWBMON_1	132	165.00	165.00	165.00	103.38
Mt Gambier	Blanche	MTGBLA_1	132	118.00	118.00	126.00	19.84
North West Bend	Robertstown 132	ROBNWB_1	132	145.00	141.00	158.00	71.80
Penola West	Kincraig	KINPEW_1	132	183.00	183.00	183.00	63.27
South East 132	Penola West	PEWSES_1	132	183.00	178.00	183.00	39.14
South East 132	Mt Gambier	SESMTG_1	132	183.00	178.00	183.00	39.14
South East 132	Snuggery	SNGSES_T	132	289.00	274.00	290.00	47.78
South West	Mt Barker South	MTBCHG_1	132	183.00	178.00	183.00	24.09
Templers	Para 132	TEMROS_1	132	137.00	137.00	137.00	9.65
Waterloo	Hummocks	HUMWAT_1	132	111.00	105.00	126.00	97.06
Waterloo	Templers	WATTEM_1	132	137.00	137.00	137.00	57.01
Waterloo	Robertstown 132	ROBMW4_2	132	137.00	137.00	137.00	27.35
Yadnarie	Port Lincoln	YADPLT_1	132	89.00	73.00	91.00	132.48
Adelaide CBD	Magill	MAGEAT_1	275	286.00	286.00	286.00	7.69
Belalie	Davenport	DVPBLI_1	275	613.00	591.00	675.00	141.52
Blyth West	Bungama 275	BUNBHW_1	275	481.00	451.00	564.00	78.87
Brinkworth 275	Davenport	DVPBRI_1	275	336.00	289.00	452.00	155.81
Bungama 275	Davenport	DVPBUN_1	275	429.00	429.00	429.00	94.83
Canowie	Mt Lock	MLOCAN_1	275	613.00	591.00	675.00	36.05
Davenport	Cultana 275	DVPCUL_2	275	457.00	457.00	457.00	66.21
Magill	North West	TOAMAG_1	275	617.00	595.00	680.00	50.66
Magill	Para 275	PARMAG_1	275	617.00	595.00	667.00	24.32
Mt Barker South	Tungkillo	MTBMB3_1	275	366.00	356.00	366.00	29.97
Mt Lock	Davenport	DVPMLO_1	275	476.00	476.00	476.00	90.03
North West	Adelaide CBD	TOBCIW_1	275	750.00	750.00	750.00	19.55
Para 275	Robertstown 275	ROBPAR_1	275	481.00	451.00	564.00	154.56
Para 275	North West	PARTOA_4	275	857.00	857.00	857.00	24.35
Para 275	Templers West	TEWPAR_1	275	336.00	289.00	452.00	1.79
Para 275	Blyth West	BHWMUN_1	275	481.00	451.00	564.00	102.48
Robertstown 275	Canowie	CANROB_1	275	476.00	476.00	476.00	83.25
Robertstown 275	Mokota	MOKROB_1	275	429.00	429.00	429.00	57.01
South West	Mt Barker South	CHGMTS_1	275	1226.00	1182.00	1350.00	23.72
South West	Magill	MAGHAV_1	275	905.00	902.00	905.00	29.72
Tailem Bend 275	South East 275	TABSES_2	275	617.00	597.00	657.00	325.90
Tailem Bend 275	South West	CHGTAB_1	275	617.00	597.00	677.00	138.81
Tailem Bend 275	Tungkillo	TUNTAB_1	275	624.00	603.00	684.00	69.30
Templers West	Brinkworth 275	BRITEW_1	275	336.00	289.00	452.00	1.79
Tungkillo	Para 275	PARTUN_1	275	962.00	896.00	1120.00	45.77
Tungkillo	Robertstown 275	ROBTUN_1	275	572.00	572.00	572.00	109.02
Willalo	Mokota	MOKWIO_1	275	613.00	591.00	675.00	12.78
Willalo	Belalie	WIOBLI_1	275	613.00	591.00	675.00	16.89
Brinkworth 132	Brinkworth 275	T_BRI	-	200.00	200.00	200.00	-
Bungama 132	Bungama 275	T_BUN	-	200.00	200.00	200.00	-
Cultana 132	Cultana 275	T_CUL	-	200.00	200.00	200.00	-
Para 132	Para 275	T_PAR		160.00	160.00	160.00	-
Robertstown 132	Robertstown 275	T_ROB	-	160.00	160.00	160.00	-
South East 132	South East 275	T_SES	-	200.00	200.00	200.00	-
Tailem Bend 132	Tailem Bend 275	T_TAB	-	160.00	160.00	160.00	-

Table A.1: Edge-list data of our simplified network. The columns Spring/Autumn, Summer and Winter are the seasonal branch capacities measured in MVA.

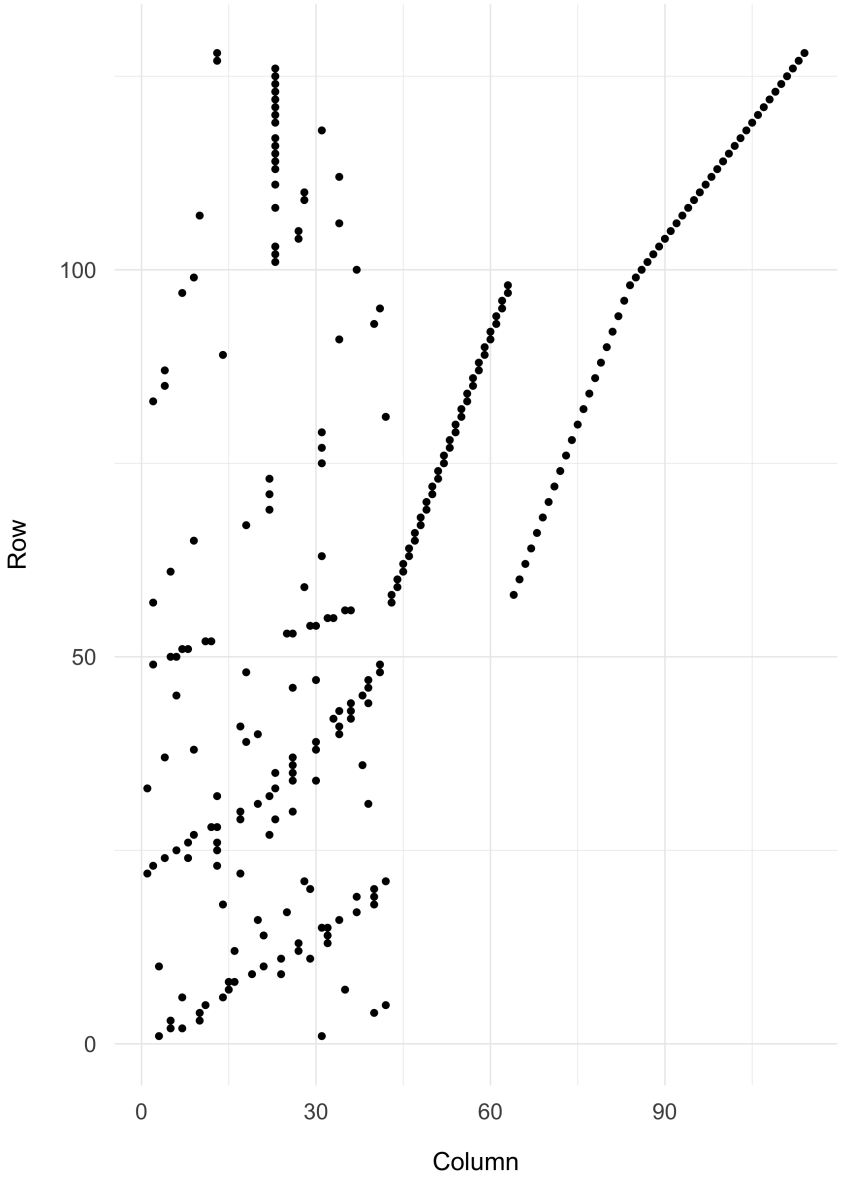


Figure A.2: The edge-vertex incidence matrix for the SA network with the additional transformer connections between the generators and high-voltage buses. The transformers correspond to the rows greater than 42 and the columns greater than 56.

Appendix B

Spectral analysis of wind generation in South Australia

The seasonality of wind generation in South Australia is explored using the power spectrum. The power spectrum is a spectral analysis tool that describes a time series by the distribution of power across the different frequencies. Significant frequencies are then used in a Fourier series for mean prediction. The effectiveness of the Fourier model relies on two components: the variance explained and the stability of the significant cycles over time. We first fit a low-pass filter in order to assess whether a Fourier model is a viable method. Then, we analyse its stability over time.

Denote by $\upsilon_k(\omega)$ the power of frequency ω in the time series of generation for wind farm k. We calculate the total power, which is also the variance explained, of a low-pass filter with cut-off corresponding to a daily cycle. The resulting Fourier series explained 78% of the variance on average. This is a high enough percentage to consider this as a possible mean prediction method. However, a low-pass fits all the frequencies that are lower than the daily cut-off and therefore over-fits the data. We now attempt to identify a smaller set of frequencies that greatly contribute to the power of the low-pass filter.

The wind power generation is high resolution data measured at 5-minute intervals which captures the variability of wind power. Higher resolution data contains more frequencies than lower resolution data and in this case the 5-minute data results in a noisy power spectrum. This makes it difficult to determine significant frequencies because it is unclear if the noise is caused by the resolution or a flat spectrum. We

propose two methods, *down-sampling* and *aggregation*, to obtain lower resolution data to determine whether the noise is a result of the high resolution data. *Down-sampling* is performed by sampling equally spaced observations of the data at a lower resolution, and *aggregation* is performed by summing the values of equally spaced intervals at a lower resolution to obtain a new aggregated series. As opposed to down-sampling, aggregation increases the signal to noise ratio to ensure the significant frequencies originate from the signal. We use down-sampling and aggregation to calculate new datasets at 30-minute, hourly and daily resolutions for the years 2016 and 2017. We rank the frequencies by total power on each dataset such that the total power of frequency ω across all the wind farms is given by

$$\widehat{\upsilon}(\omega) = \frac{1}{n}\sum_{k=1}^{n}\frac{\upsilon_k(\omega)}{\sum_{\omega}\upsilon_k(\omega)}$$

where $\upsilon_k(\omega)$ is the power spectrum of wind farm k and n is the number of wind farms. We rank the frequencies on their total power, $\widehat{\upsilon}(\omega)$, in Table B.1.

	5-min	30-min	60-min	daily
1	0.97	0.97	0.97	0.97
2	365.50	365.50	365.50	4.87
3	2.44	2.44	2.44	2.44
4	4.87	4.87	4.87	5.85
5	64.82	64.82	29.73	30.70
6	29.73	29.73	64.82	77.00
7	30.70	30.70	30.70	64.82
8	77.00	77.00	77.00	24.37
9	39.47	39.47	39.47	37.52
10	1.95	1.95	1.95	53.61
11	37.52	37.52	37.52	39.47
12	19.01	19.01	19.01	26.32
13	72.61	72.61	72.61	19.01
14	1.46	1.46	1.46	1.46
15	24.37	24.37	24.37	107.21
16	5.85	5.85	5.85	8.77
17	36.55	36.55	36.55	29.73
18	101.85	101.85	101.85	46.78
19	58.97	58.97	58.97	1.95
20	56.04	56.04	56.04	72.61

(a) Down-sampled.

	5-min	30-min	60-min	daily
1	0.97	0.97	0.97	0.97
2	365.50	365.50	365.50	2.44
3	2.44	2.44	2.44	4.87
4	4.87	4.87	4.87	29.73
5	64.82	64.82	64.82	30.70
6	29.73	29.73	29.73	64.82
7	30.70	30.70	30.70	39.47
8	77.00	77.00	77.00	1.95
9	39.47	39.47	39.47	77.00
10	1.95	1.95	1.95	37.52
11	37.52	37.52	37.52	72.61
12	19.01	19.01	19.01	19.01
13	72.61	72.61	72.61	1.46
14	1.46	1.46	1.46	24.37
15	24.37	24.37	24.37	5.85
16	5.85	5.85	5.85	36.55
17	36.55	36.55	36.55	101.85
18	101.85	101.85	101.85	58.97
19	58.97	58.97	58.97	53.61
20	56.04	56.04	56.04	56.04

(b) Aggregated.

Table B.1: Ranking of the frequencies with the highest total power, $\widehat{\upsilon}(\omega)$, across all wind farms for each dataset. The columns 5-min is the significant cycles of the original data.

The results from Table B.1 demonstrate consistency amongst the most significant frequencies. We fit a Fourier series corresponding to the significant cycles above for the years 2016-2018 and record the corresponding variance explained in Table B.2. This demonstrates that the Fourier series explains only at most 24% of the vari-

ance, which decreases to 17% in 2018. Therefore, we conclude that the spectrum is relatively flat and that cycles of wind generation may change from year to year.

	2016	2017	2018
BLUFF1	0.22	0.18	0.16
CATHROCK	0.13	0.17	0.11
CLEMGPWF	0.14	0.14	0.10
CNUNDAWF	0.17	0.17	0.17
HALLWF1	0.24	0.17	0.17
HALLWF2	0.23	0.18	0.15
LKBONNY1	0.17	0.17	0.15
LKBONNY2	0.18	0.16	0.13
LKBONNY3	0.18	0.16	0.13
MTMILLAR	0.17	0.23	0.13
NBHWF1	0.21	0.16	0.15
SNOWNTH1	0.17	0.17	0.12
SNOWSTH1	0.18	0.17	0.11
SNOWTWN1	0.14	0.17	0.13
STARHLWF	0.18	0.18	0.15
WATERLWF	0.21	0.18	0.16
WPWF	0.13	0.17	0.13

Table B.2: Variance explained by the selected cycles for each year.

Appendix C

Processing maintenance

The time series of generation data for 2015 from AEMO [45] contains periods where the power output of a wind farm appears constrained. We believe this is caused by either an electrical feeder malfunction or general maintenance. Since this is a large dataset with over 100,000 observations for each wind farm, manually removing these points would be too time consuming. Therefore, we apply a basic set of rules in order to identify the maintenance times within the time series of generation for each wind farm. Maintenance corresponds to a point such that

- if a wind farm has no power output and the average of all other wind farms within the zone is greater than the cuttoff c, or
- if a wind farm has power output $p\%$ less than the average of all other wind farms within the zone for at least h hours.

We use different values for the parameters c, p and h, as the within-zone correlation was different for each zone. Below we outline the values of the parameters.

Zone	c	p	h
South East	0.1 pu	20%	1 hour
Mid-North (East)	0.2 pu	40%	3 hours
Mid-North (West)	0.2 pu	30%	2 hours

Table C.1: Maintenance processing parameter values.

We recognise that this processing method might not be perfect; however, superior methods will have negligible effect on the parameter estimation due to the size of the dataset.

Appendix D

Diagnostic plots for the wind power simulation

In Chapter 5, the goodness-of-fit of the mean prediction model and error simulation model is assessed using various diagnostic plots. These are produced by applying a *leave-one-out* technique and treating the power output of the removed wind farm as unknown. We do this to emulate predicting values of a new wind farm whose power output is unknown. Within Chapter 5, we present the diagnostic plots for the case where the Snowtown North Wind Farm (SNOWNTH1) is treated as unknown. The diagnostic plots corresponding to all other wind farms are presented here.

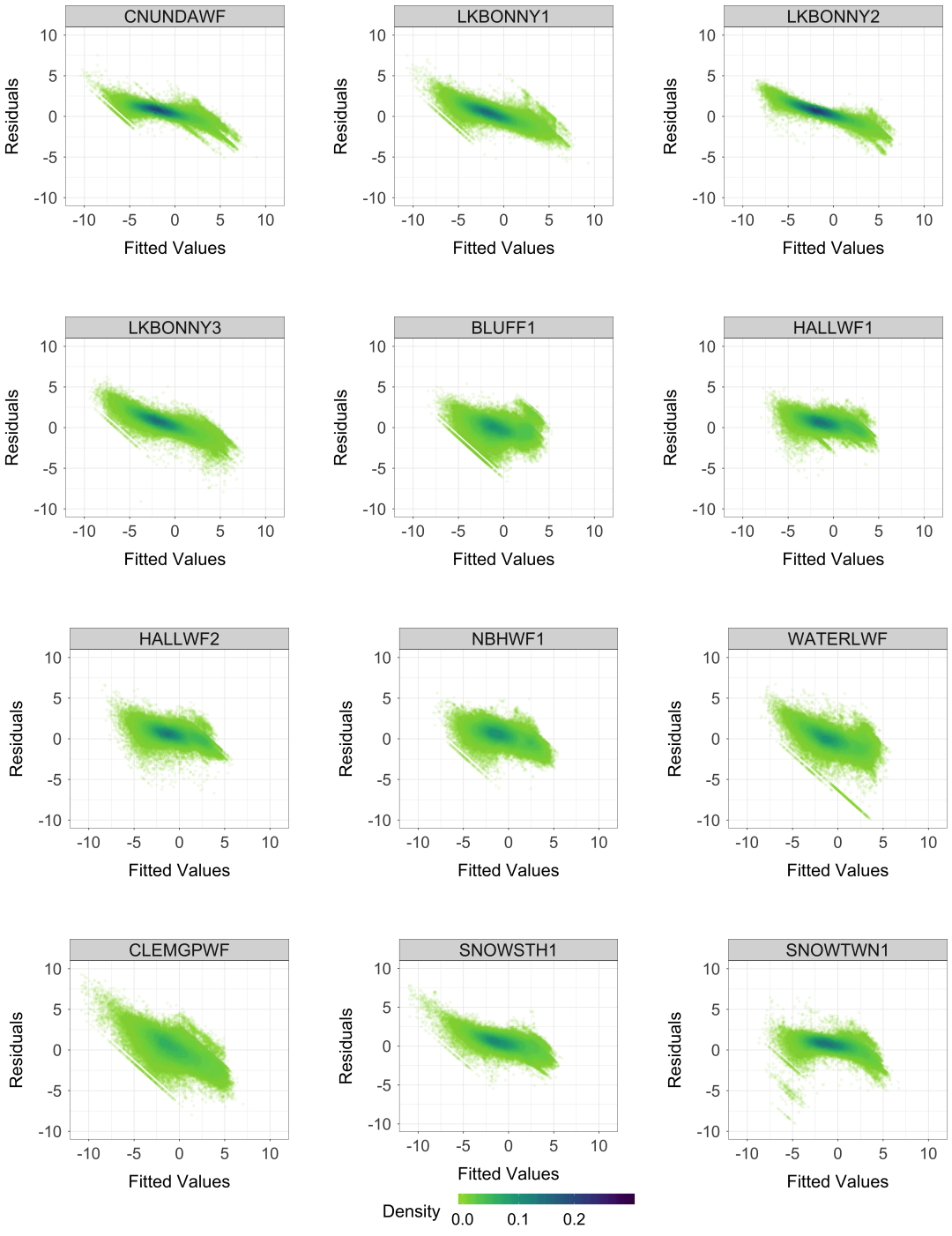


Figure D.1: Residual versus fitted values of the average regression model. The x-axis is the fitted values, η_t^*, and the y-axis is the difference between the actual values and the fitted values. A linear trend-line is evident in the residuals.

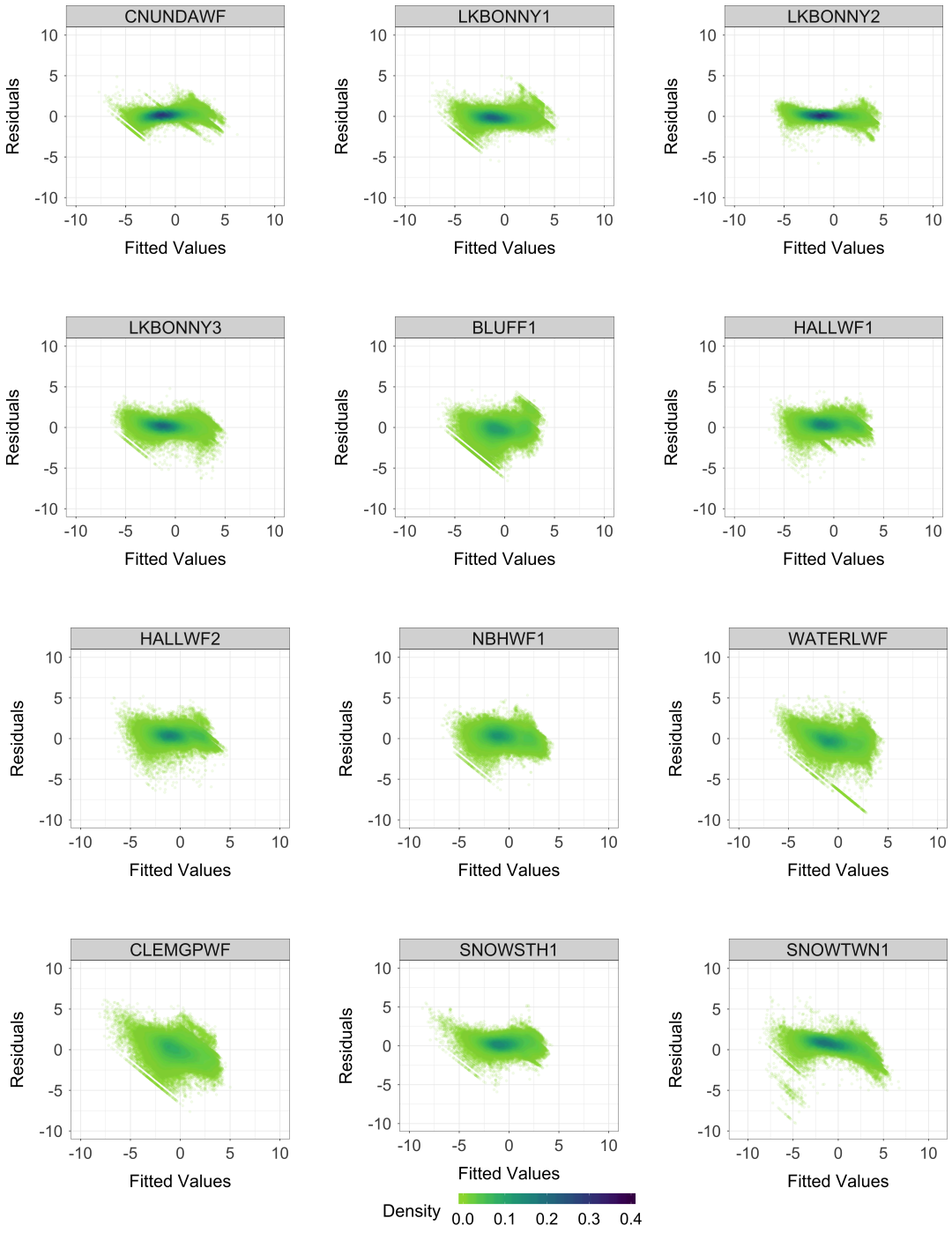


Figure D.2: Residual versus fitted values of the corrected average regression model. The x-axis is the fitted values, $\widehat{\eta}_t^*$, and the y-axis is the difference between the actual values and the fitted values. The average correlation coefficient of -0.08 suggests a relatively small linear trend.

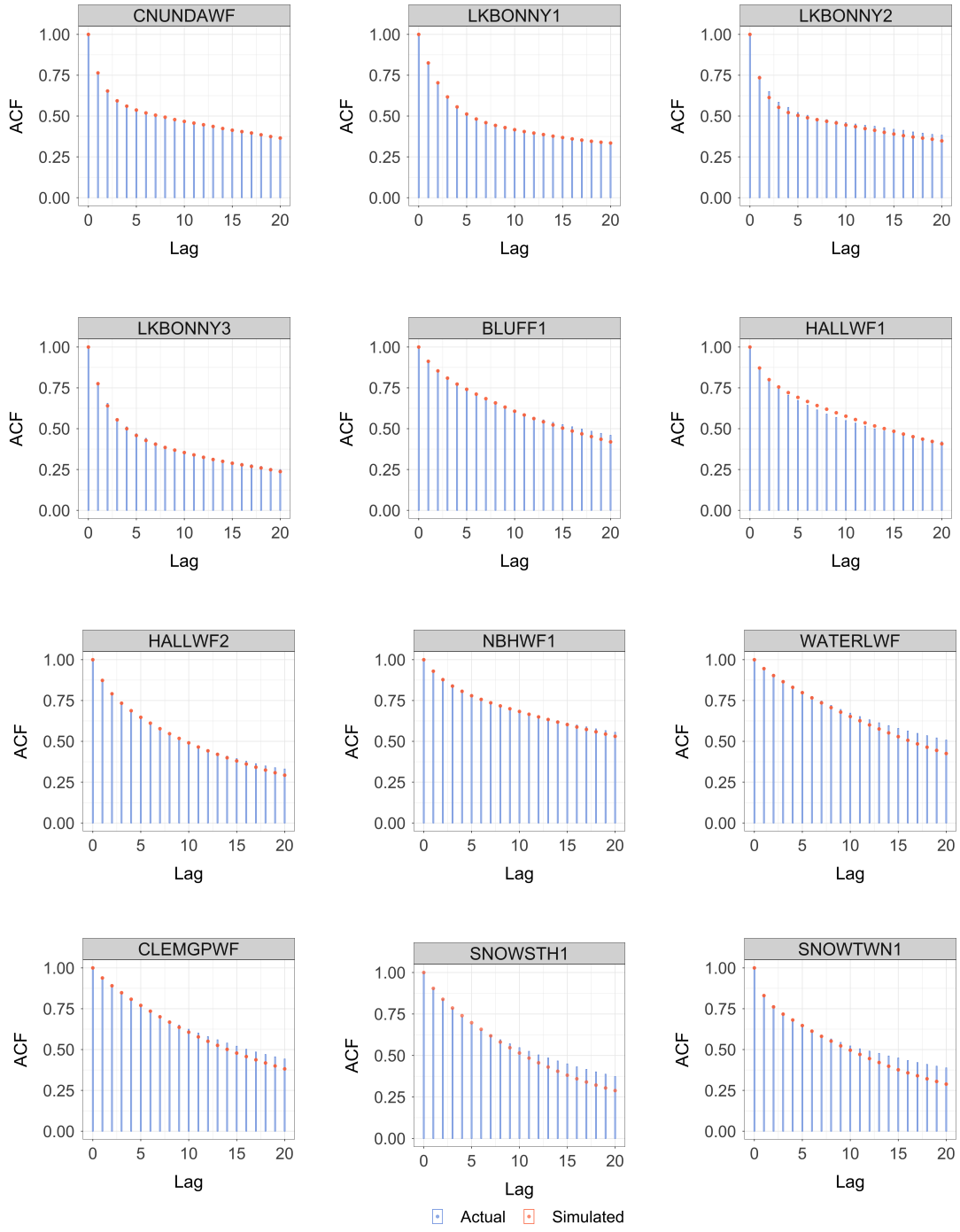


Figure D.3: Auto-correlation of the actual residuals and the simulated residuals, ζ_t^*. Simulations are of an ARMA(2,1) time series model fitted to the residuals of the corrected average regression model. The simulations of the ARMA(2,1) model the strong auto-correlation observed in the residuals.

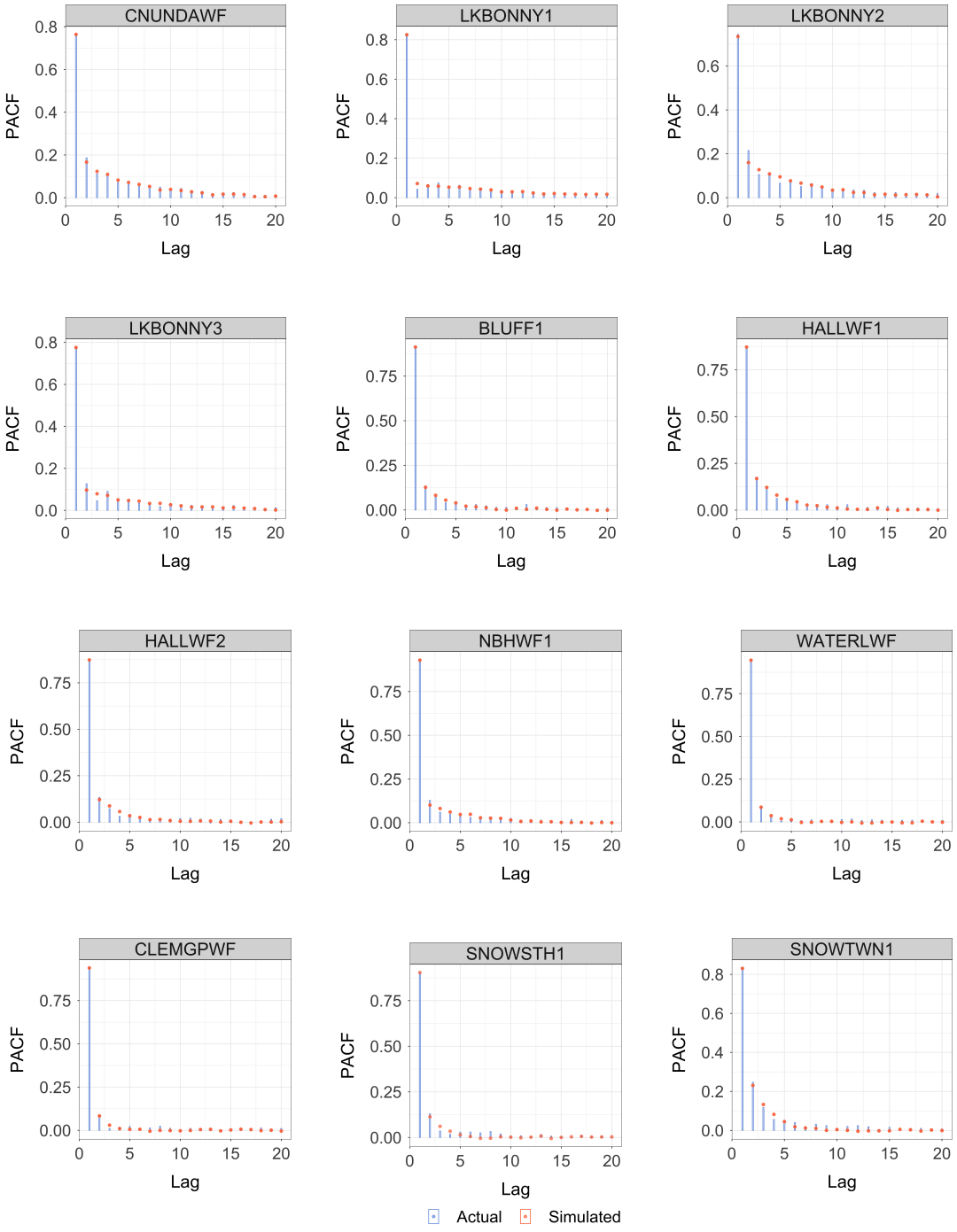


Figure D.4: Partial auto-correlation of the actual residuals and the simulated residuals, ζ_t^*. Simulations are of an ARMA(2,1) time series model fitted to the residuals of the corrected average regression model. The simulations of the ARMA(2,1) accurately model the partial auto-correlation.

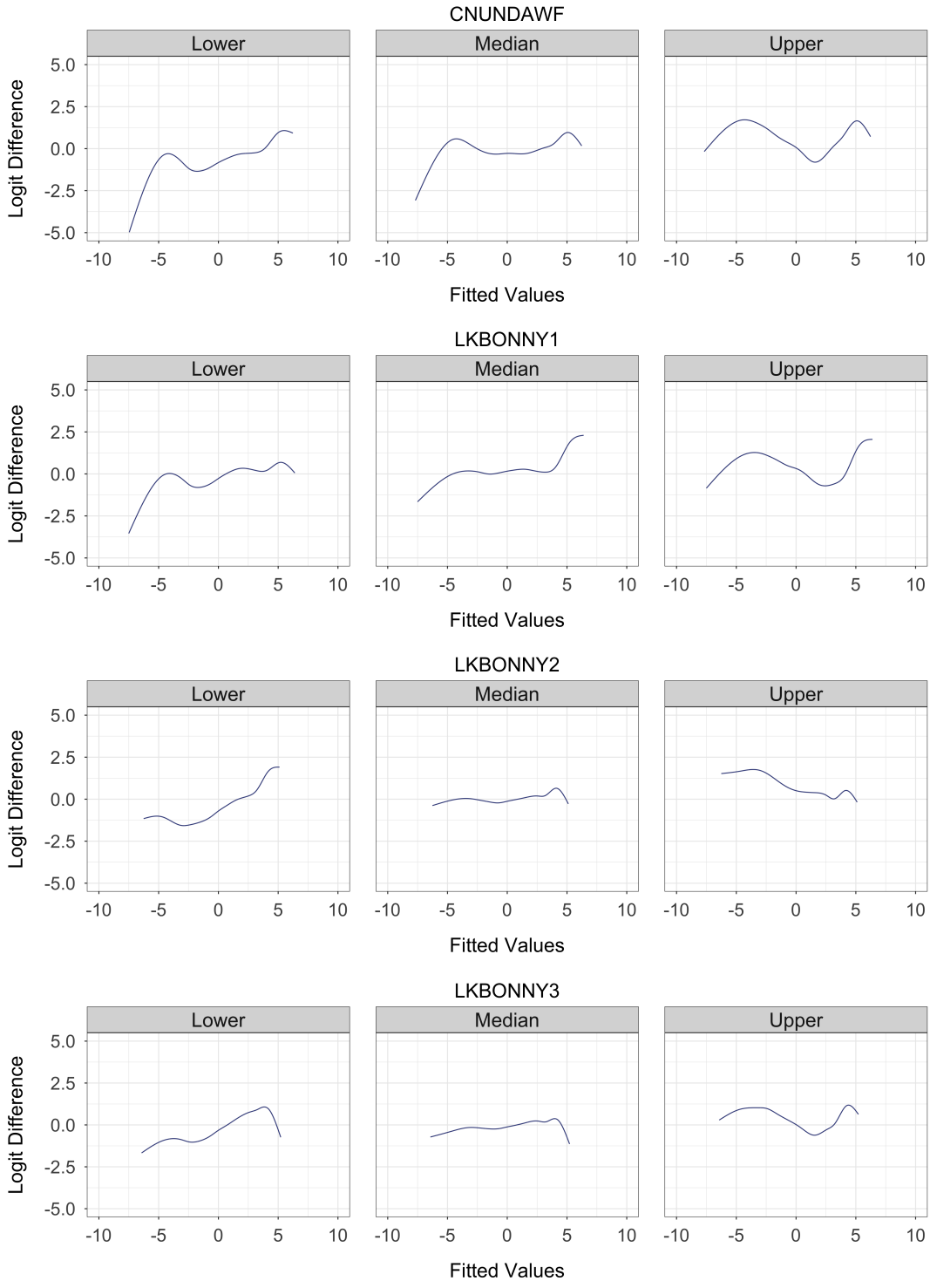


Figure D.5: The fit of the simulated residuals compared to the actual residuals when each wind farm is treated as unknown. The x-axis is the fitted values, $\widehat{\eta}_t^*$, of the corrected average regression. The y-axis is the logit difference between the averaged quantile function, $Q_{\widehat{e}_t^*|\widehat{\eta}_t^*}(\tau)$, and the actual residual quantiles evaluated at $\tau = 0.05, 0.5, 0.95$. The method tends to underestimate the lower and median quantiles when the expected wind power output is low.

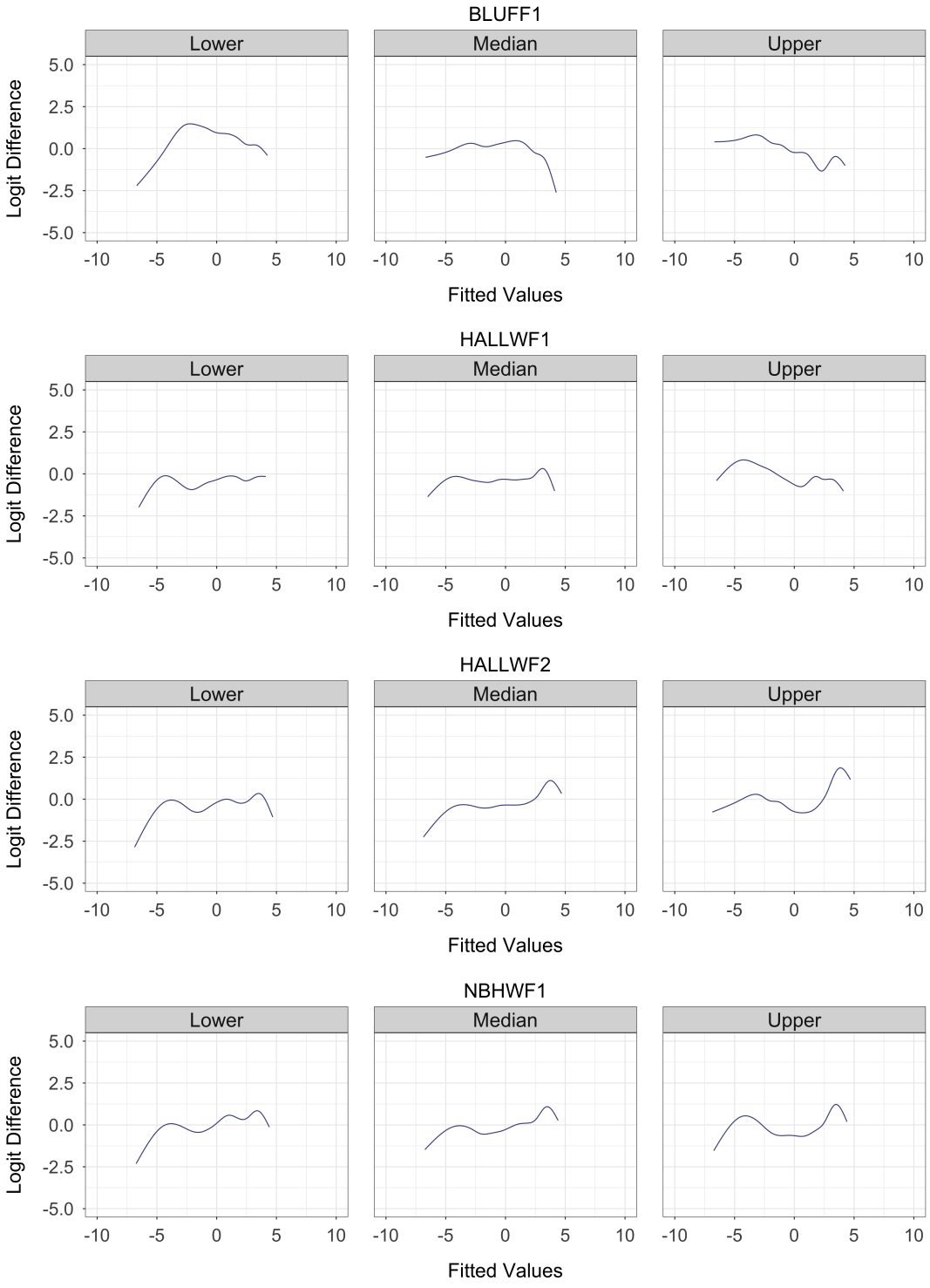


Figure D.5 (cont.): The fit of the simulated residuals compared to the actual residuals when each wind farm is treated as unknown. The x-axis is the fitted values, $\widehat{\eta}_t^*$, of the corrected average regression. The y-axis is the logit difference between the averaged quantile function, $Q_{\widehat{e}_t^*|\widehat{\eta}_t^*}(\tau)$, and the actual residual quantiles evaluated at $\tau = 0.05, 0.5, 0.95$. The method tends to underestimate the lower and median quantiles when the expected wind power output is low.

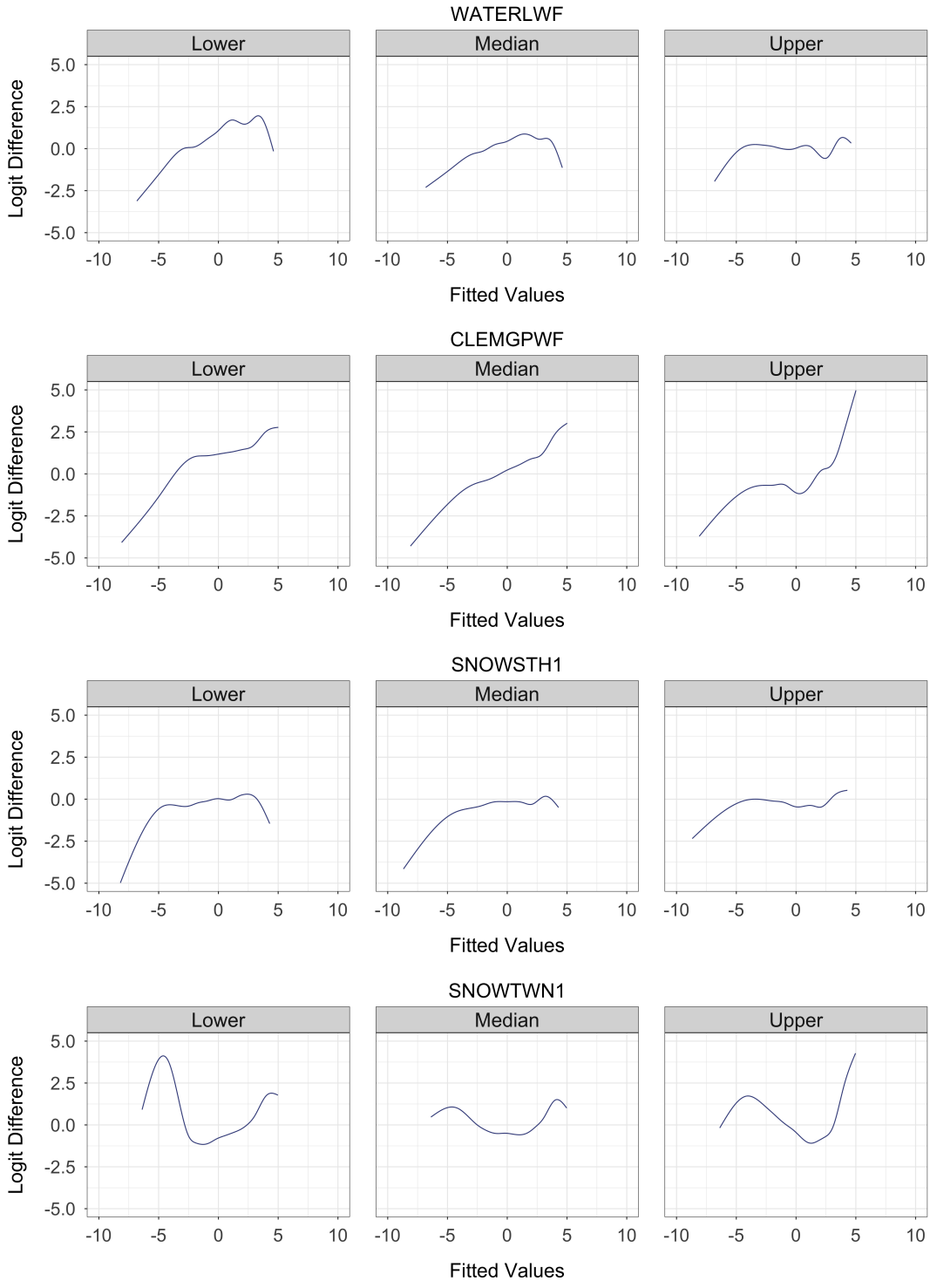


Figure D.5 (cont.): The fit of the simulated residuals compared to the actual residuals when each wind farm is treated as unknown. The x-axis is the fitted values, $\widehat{\eta}_t^*$, of the corrected average regression. The y-axis is the logit difference between the averaged quantile function, $Q_{\widehat{e}_t^*|\widehat{\eta}_t^*}(\tau)$, and the actual residual quantiles evaluated at $\tau = 0.05, 0.5, 0.95$. The method tends to underestimate the lower and median quantiles when the expected wind power output is low.

Bibliography

[1] Geoscience Australia. Electricity transmission lines. https://data.gov.au/dataset/electricity-transmission-lines, 2017. [Online; accessed 15 Apr. 2018].

[2] REpower Australia. Wind farm developments in South Australia: Select committee inquiry. https://www.senvion.com/fileadmin/Redakteur/Hubs/Austral-Asia/Submissions/SA_Select_Committee_Inquiry_Final.pdf, 2020. [Online; accessed 8 Apr. 2020].

[3] Department of Energy Australian Government. Australian energy statistics. https://www.energy.gov.au/publications/australian-energy-statistics-table-o-electricity-generation-fuel-type-2017-18-and-2018, 2019. [Online; accessed 12 Feb. 2020].

[4] Andrey Bernstein, Daniel Bienstock, David Hay, Meric Uzunoglu, and Gil Zussman. Sensitivity analysis of the power grid vulnerability to large-scale cascading failures. *ACM SIGMETRICS Performance Evaluation Review*, 40(3):33–37, 2012.

[5] Daniel Bienstock. *Electrical transmission system cascades and vulnerability: an operations research viewpoint*, volume 22. SIAM, 2015.

[6] Daniel Bienstock and Abhinav Verma. The $N-k$ problem in power grids: New models, formulations, and numerical experiments. *SIAM Journal on Optimization*, 20(5):2352–2380, 2010.

[7] Lori Bird, Debra Lew, Michael Milligan, Enrico Carlini, Ana Estanqueiro, Damian Flynn, Emilio Gomez-Lazaro, Hannele Holttinen, Nickie Menemenlis, Antje Orths, et al. Wind and solar energy curtailment: A review of international experience. *Renewable and Sustainable Energy Reviews*, 65:577–586, 2016.

[8] Lori Bird, Jaquelin Cochran, and Xi Wang. Wind and solar energy: experience and practices in the United States. https://www.nrel.gov/docs/fy14osti/60983.pdf, 2018. [Online; accessed 15 May 2020].

[9] John Boland. Time series modelling of solar radiation. In *Modeling Solar Radiation at the Earth's Surface*, pages 283–312. Springer, 2008.

[10] John Boland and Adrian Grantham. Nonparametric conditional heteroscedastic hourly probabilistic forecasting of solar radiation. *Multidisciplinary Scientific Journal*, 1(1):174–191, 2018.

[11] Ettore Bompard, Pedro Correia, George Gross, and Mikael Amelin. Congestion-management schemes: a comparative analysis under a unified framework. *IEEE transactions on power systems*, 18(1):346–352, 2003.

[12] Daniel Burke and Mark O'Malley. Factors influencing wind energy curtailment. *IEEE Transactions on Sustainable Energy*, 2(2):185–193, 2011.

[13] Hale Cetinay, Fernando Kuipers, and Piet Van Mieghem. A topological investigation of power flow. *IEEE Systems Journal*, pp:1–9, 2017.

[14] Dhiman Chatterjee, Jeff Webb, Qun Gao, Michael Vaiman, Marianna Vaiman, and Mark Povolotskiy. N-1-1 AC contingency analysis as a part of NERC compliance studies at midwest ISO. In *IEEE PES T&D 2010*, pages 1–7. IEEE, 2010.

[15] Ian Dobson, Benjamin Carreras, Vickie Lynch, and David Newman. An initial model for complex dynamics in electric power system blackouts. In *the 34th Hawaii International Conference on System Sciences*. IEEE, January 2001.

[16] Ian Dobson, Benjamin Carreras, and David Newman. A probabilistic loading-dependent model of cascading failure and possible implications for blackouts. In *System Sciences, 2003. Proceedings of the 36th Annual Hawaii International Conference on*, pages 10–pp. IEEE, 2003.

[17] Ian Dobson, Benjamin Carreras, and David Newman. A branching process approximation to cascading load-dependent system failure. In *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference*. IEEE, 2004.

[18] Ian Dobson, Benjamin Carreras, Vickie Lynch, and David Newman. Complex systems analysis of series of blackouts: Cascading failure, critical points, and

self-organization. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 17(2):026103, 2007.

[19] Ian Dobson, Benjamin Carreras, David Newman, and José Reynolds-Barredo. Obtaining statistics of cascading line outages spreading in an electric transmission network from standard utility data. *IEEE Transactions on Power Systems*, 31(6):4831–4841, 2016.

[20] ElectraNet. Power system strength. https://www.electranet.com.au/wp-content/uploads/2018/05/180103_ElectraNet_System-Strength_Fact-Sheet_WEB.pdf, 2018. [Online; accessed 10 Apr. 2020].

[21] DP Energy. Port Augusta renewable energy park. http://dpenergy.info/parep/the-project, 2019. [Online; accessed 8 Apr. 2020].

[22] Jean-Baptiste Fourier. *Théorie analytique de la chaleur*. Firmon Didot, 1822.

[23] Yingzhong Gu and Le Xie. Fast sensitivity analysis approach to assessing congestion induced wind curtailment. *IEEE Transactions on Power Systems*, 29(1):101–110, 2013.

[24] Yingzhong Gu, Le Xie, Brett Rollow, and Bo Hesselbaek. Congestion-induced wind curtailment: Sensitivity analysis and case studies. In *2011 North American Power Symposium*, pages 1–7. IEEE, 2011.

[25] James Hamilton. *Time series analysis*, volume 2. Princeton New Jersey, 1994.

[26] Bálint Hartmann, István Vokony, and István Táczi. Effects of decreasing synchronous inertia on power system dynamics—overview of recent experiences and marketisation of services. *International Transactions on Electrical Energy Systems*, 29(12):e12128, 2019.

[27] William Horrace. Some results on the multivariate truncated normal distribution. *Journal of multivariate analysis*, 94(1):209–221, 2005.

[28] Nick Jenkins and Goran Strbac. *Electric power systems*. John Wiley and Sons, Ltd., Chichester, West Sussex, UK, 5th ed. edition, 2012. ISBN 1-118-36110-5.

[29] Michael Joos and Iain Staffell. Short-term integration costs of variable renewable energy: Wind curtailment and balancing in Britain and Germany. *Renewable and Sustainable Energy Reviews*, 86:45–65, 2018.

[30] Marek Kimmel and David Axelrod. *The Galton-Watson Process*, pages 33–63. Springer New York, New York, NY, 2002.

[31] Roger Koenker and Pin Ng. Inequality constrained quantile regression. *Sankhyā: The Indian Journal of Statistics*, 67:418–440, 2005.

[32] Ashwani Kumar, Suresh Srivastava, and Sri Singh. Congestion management in competitive power market: A bibliographical survey. *Electric Power Systems Research*, 76(1-3):153–164, 2005.

[33] Guo-liang Luo, Yan-ling Li, Wen-jun Tang, and Xiao Wei. Wind curtailment of China's wind power operation: Evolution, causes and solutions. *Renewable and Sustainable Energy Reviews*, 53:1190–1201, 2016.

[34] Edward McGarrigle, John Deane, and Paul Leahy. How much wind energy will be curtailed on the 2020 Irish power system? *Renewable Energy*, 55:544–553, 2013.

[35] Eoghan McKenna, Philipp Grünewald, and Murray Thomson. Going with the wind: temporal characteristics of potential wind curtailment in Ireland in 2020 and opportunities for demand response. *IET Renewable Power Generation*, 9 (1):66–77, 2015.

[36] Carl Meyer. Generalized inversion of modified matrices. *SIAM Journal on Applied Mathematics*, 24(3):315–323, 1973.

[37] Debasish Mondal, Aparajita Sengupta, and Abhijit Chakrabarti. Chapter 7 - Application of FACTS controller. In Debasish Mondal, Abhijit Chakrabarti, and Aparajita Sengupta, editors, *Power System Small Signal Stability Analysis and Control*, pages 185 – 225. Academic Press, Boston, 2014. ISBN 978-0-12-800572-9.

[38] Tommaso Nesti, Alessandro Zocca, and Bert Zwart. Emergent failures and cascades in power grids: a statistical physics perspective. *Physical Review Letters*, 120(25):258–301, 2018.

[39] South Australian Power Networks. Zone substation data. https://www.sapowernetworks.com.au/data/, 2019. [Online; accessed 11 Aug. 2019].

[40] South Australian Power Networks. Excluded substations statement. https://www.sapowernetworks.com.au/data/3034/excluded-substations-statement/, 2019. [Online; accessed 11 Aug. 2019].

[41] Bureau of Meteorology. Weather station directory. http://www.bom.gov.au/climate/data/stations/, 2019. [Online; accessed 3 May 2019].

[42] Department of the Environment and Energy. Electricity generation by fuel type. https://www.energy.gov.au/publications/australian-energy-statistics-table-o-electricity-generation-fuel-type-2017-18-and-2018, 2019. [Online; accessed 27 Feb. 2020].

[43] Australian Energy Market Operator. Black System South Australia 28 September 2016. https://www.aemo.com.au/-/media/Files/Electricity/NEM/Market_Notices_and_Events/Power_System_Incident_Reports/2017/Integrated-Final-Report-SA-Black-System-28-September-2016.pdf, 2017. [Online; accessed 13 Feb. 2020].

[44] Australian Energy Market Operator. South Australia system strength assessment. https://cdn.theconversation.com/static_files/files/46/SA_System_Strength_201720180326-22189-lnq050.pdf?1522072786, 2017. [Online; accessed 10 Feb. 2020].

[45] Australian Energy Market Operator. Generation and load. https://www.aemo.com.au/Electricity/National-Electricity-Market-NEM/Data/Market-Management-System-MMS/Generation-and-Load, 2018. [Online; 13 Oct. 2018].

[46] Australian Energy Market Operator. Transmission equipment ratings. https://www.aemo.com.au/Electricity/National-Electricity-Market-NEM/Data/Network-Data/Transmission-Equipment-Ratings, 2018. [Online; accessed 3 Sep. 2019].

[47] Australian Energy Market Operator. Integrated System Plan 2018. https://www.aemo.com.au/-/media/Files/Electricity/NEM/Planning_and_Forecasting/ISP/2018/Integrated-System-Plan-2018_final.pdf, 2018. [Online; 10 Feb. 2020].

[48] Australian Energy Market Operator. Quarterly energy dynamics - Q3 2018. https://aemo.com.au/-/media/files/major-publications/qed/2018/qed-q3-2018.pdf?la=en, 2018. [Online; accessed 16 Oct. 2020].

[49] Australian Energy Market Operator. South Australia electricity report 2017. https://www.aemo.com.au/-/media/Files/Electricity/NEM/Planning_and_Forecasting/SA_Advisory/2017/South-Australian-Electricity-Report-2017.pdf, 2018. [Online; accessed 10 Feb. 2020].

[50] Australian Energy Market Operator. Market data. https://www.nemweb.com.au/#mms-data-model, 2019. [Online; accessed 20 May 2019].

[51] Australian Energy Market Operator. Generation information. https://aemo.com.au/en/energy-systems/electricity/national-electricity-market-nem/nem-forecasting-and-planning/forecasting-and-planning-data/generation-information, 2019. [Online; 13 Oct. 2018].

[52] Australian Energy Market Operator. South Australia electricity report 2019. https://www.aemo.com.au/-/media/Files/Electricity/NEM/Planning_and_Forecasting/SA_Advisory/2019/2019-South-Australian-Electricity-Report.pdf, 2019. [Online; accessed 10 Feb. 2020].

[53] Australian Energy Market Operator. Diagrams and previous maps. https://aemo.com.au/en/energy-systems/electricity/national-electricity-market-nem/nem-forecasting-and-planning/forecasting-and-planning-data/diagrams-and-previous-maps, 2020. [Online; accessed 17 May 2019].

[54] Konrad Purchala, Leonardo Meeus, Daniel Van Dommelen, and Ronnie Belmans. Usefulness of DC power flow for active power flow analysis. In *Power Engineering Society General Meeting, 2005. IEEE*, pages 454–459. IEEE, 2005.

[55] Junjian Qi, Ian Dobson, and Shengwei Mei. Towards estimating the statistics of simulated cascades of outages with branching processes. *IEEE Transactions on Power Systems*, 28(3):3410–3419, 2013.

[56] Tilt Renewables. Palmer wind farm. https://www.tiltrenewables.com/assets-and-projects/Palmer-Wind-Farm/, 2020. [Online; accessed 8 Apr. 2020].

[57] TSCNET Services. ENTSO-E report on frequency drop. https://www.tscnet.eu/tag/frequency/, 2020. [Online; accessed 12 June. 2020].

[58] Saleh Soltan, Dorian Mazauric, and Gil Zussman. Analysis of failures in power grids. *IEEE Transactions on Control of Network Systems*, 4(2):288–300, 2017.

[59] Shengpeng Sun, Fengliang Liu, Song Xue, Ming Zeng, and Fanxiao Zeng. Review on wind power development in China: current situation and improvement strategies to realize future development. *Renewable and Sustainable Energy Reviews*, 45:589–599, 2015.

[60] Renewable Energy Systems. Twin creek wind farm. http://www.twincreek-windfarm.com/news/latest-project-news/, 2020. [Online; accessed 8 Apr. 2020].

[61] Hugo Touchette. The large deviation approach to statistical mechanics. *Physics Reports*, 478(1):1–69, 2009.

[62] James Turner, Darryl Downing, and James Bogard. Regression analysis. In *Statistical Methods in Radiation Physics*, pages 353–385. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 2012. ISBN 9783527646548.

[63] Kenneth Van den Bergh, Dries Couckuyt, Erik Delarue, and William D'haeseleer. Redispatching in an interconnected electricity system with high renewables penetration. *Electric Power Systems Research*, 127:64–72, 2015.

[64] Zhifang Wang, Anna Scaglione, and Robert Thomas. A Markov-transition model for cascading failures in power grids. In *2012 45th Hawaii International Conference on System Sciences*, pages 2115–2124. IEEE, 2012.

[65] WindEurope. Wind energy in Europe 2018. https://windeurope.org/wp-content/uploads/files/about-wind/statistics/WindEurope-Annual-Statistics-2018.pdf, 2019. [Online; accessed 17 Mar. 2020].

[66] Jerrold Zar. Significance testing of the spearman rank correlation coefficient. *Journal of the American Statistical Association*, 67(339):578–580, 1972.

[67] Ray Daniel Zimmerman and Carlos Edmundo Murillo-Sánchez. Matpower (version 7.0). https://matpower.org, 2019. [Software].

[68] Ray Daniel Zimmerman, Carlos Edmundo Murillo-Sánchez, and Robert John Thomas. Matpower: Steady-state operations, planning, and analysis tools for power systems research and education. *IEEE Transactions on power systems*, 26(1):12–19, 2010.