THE UNIVERSITY
of EDINBURGH

Power system adequacy: on two-area models and the capacity procurement decision process

Nestor Sanchez Guadarrama

Doctor of Philosophy
University of Edinburgh
2022

Declaration

I declare that this thesis was composed by myself and that the work contained therein is my own, except where explicitly stated otherwise in the text.

(*Nestor Sanchez Guadarrama*)

Lay summary

A reliable supply of electricity is a critical need for all modern societies. For this reason, risk quantification in the electricity supply has been an active area of research almost since the start of electrification, and is referred to as power system adequacy. Its main goal is aiding decision makers in long term system planning. Recent technological and economic developments in power systems, such as the large scale introduction of renewable technologies like wind and solar power in the system's generation mix, as well as the proliferation of interconnectors to neighbouring power systems, have introduced new challenges in system adequacy modelling, or made existing challenges more pressing. Appropriate risk quantification models that account for these transformations are crucial to make the transition to a low carbon power system without compromising consumers' access to electricity or putting economic growth at risk.

In this work, we mainly explore two problems. One is devising appropriate statistical models for the occurrence of relevant extreme events in the system (at least one of very high demand or very low available capacity, e.g. lower than expected wind generation); this is important because almost all of the risks to electricity supply come from extreme events, even though system adequacy models often do not account for this explicitly. We explore said models in an interconnected two-area systems, using results from the statistical theory of extreme values. We find that these models can produce material differences in risk estimates when compared to existing adequacy models.

The second one is analysing the effects of installing large amounts of wind generation capacity in a system like the one in Great Britain. While adequacy studies are usually limited to the calculation of unidimensional risk indices based on long-term averages (i.e. expected values) of the statistics of interest, we perform a much deeper simulation analysis of loss of load event sizes, durations and spread. We find that the usual expected value risk indices become progressively insufficient to convey a useful summary of system risk as the amount of installed wind capacity grows. In other words, the typical indices that have been used for

many decades now to measure adequacy risks in power systems may not convey a meaningful picture of system risk when renewable generation represents a substantial part of the generation mix. This is because the additional variation that is introduced in the system due to the reliance on highly variable weather patterns, make long-term average indices unrepresentative of year to year adequacy risks.

Abstract

In this work, we explore methodological extensions to modelling practices in power system adequacy for single-area and two-area systems. Specifically, we build on top of some of the practices currently in use in Great Britain (GB) by National Grid, framing this in the context of the current technological transition in which renewable capacity is gradually replacing a considerable share of fossil-fuel-based capacity.

We explore two-area extensions of the methodology currently used in GB to quantify risk in single-area models. By doing this, we also explore the impact of shortfall-sharing policies and wind capacity on risk indices and on the value of interconnection. Furthermore, we propose a model based on the statistical theory of extreme values to characterise statistical dependence across systems in both net demand (defined as power demand minus renewable generation) and capacity surpluses/deficits (defined as power supply minus demand), looking at how statistical dependence strength influences post-interconnection risk and the capacity value of interconnection. Lastly, we analyse the risk profile of a single-area system as reliance on wind capacity grows, looking at risk beyond the standard set of risk indices, which are based on long-term averages. In doing this, we look at trends which are overlooked by the latter, yet are of considerable importance for decision-makers. Moreover, we incorporate a measure of the decision-maker's degree of risk aversion into the current capacity procurement methodology in GB, and look at the impact of this and other parameters on the amount of procured capacity.

We find that shortfall-sharing policies can have a sizeable impact on the interconnector's valuation in terms of security of supply, specially for systems that are significantly smaller than their neighbours. Moreover, this valuation also depends strongly on the risk indices chosen to measure it. We also find that the smoothing effect of parametric extreme value models on tail regions can have a material effect on practical adequacy calculations for post-interconnection risks, and that assumed independence between conventional generation fleets makes

capacity shortfall co-occurrences only weakly dependent (in a precisely defined sense) across areas despite much stronger statistical dependence between system net demands. Lastly, as more wind capacity is installed, we find multiple relevant changes in the (single-area) system's risk profile that are not expressed by the standard risk indices: in particular, we find a substantial increase in the frequency of severe events, extreme year-to-year variability of outturn, and a progression to a system with fewer days of potentially much larger shortfalls. Moreover, we show that a high reliance on wind introduces a substantial amount of uncertainty into the calculations due to the limited number of available historic years, which cannot account for the wide range of possible weather conditions the system could experience in the future. Lastly, we also find that the a higher reliance on wind generation also impact the capacity procurement decision process, potentially making the amount of procured capacity considerably more sensitive to parameters such as the value of lost load.

Acknowledgements

I am grateful to both of my supervisors, Chris J. Dent and Amy Wilson, for their outstanding guidance and support throughout this project, to the Mexican Council of Science and Technology for their financial support, and to Ioannis Papastathopoulos and Stan Zachary for the many insightful discussions. I am also grateful to all of the friends I made along the way, and most of all to Magda, for making these years so colourful and for being with me in the ups and downs.

Contents

Chapter 1

Introduction

This Chapter outlines the context in which this thesis is framed. It discusses the current transformation power systems are going through, and motivates the research on power system adequacy. Furthermore, it delimits and motivates the specific challenges this work intends to address, and outlines the outputs of this research project, which go beyond this thesis. Lastly, the organisation of this thesis is explained.

1.1 Problem's context

Reliability, along with affordability and environmental sustainability, is one of the pillars of energy policy in Great Britain (GB) and many other power systems around the world. In this context, reliability is understood as the ability of the system to function in a steady state for a sustained period of time. Power grids are complex systems that can cover entire geographical regions and have a very large number of individual components. For this reason reliability analyses tend to focus on individual aspects of the system separately; in general we can make a distinction between three main subsystems: generation, transmission and distribution. The generation subsystem is comprised of all generating power plants, while the transmission system moves generated electricity over long distances to population and industrial hubs, and distribution networks deliver electricity to individual customers.

Within the study of power system reliability, in this work we focus on the problem of *resource adequacy*, this is, the existence of sufficient generation capacity to satisfy electricity demand, which is mainly to do with the generation subsystem (Billinton and Allan, 1984). The purpose of this kind of analysis is to aid decision-makers and operators in capacity expansion planning for future

years, ensuring security of supply standards are met despite changes in demand patterns, the composition of the generation mix and other factors. Since in this context the main concern are decisions on planning time scales of multiple years, adequacy is usually analysed separately from *security*, which is understood as the ability of the system to withstand unforeseen disturbances such as the loss of transmission facilities, and which is associated to operational timescales, i.e., closer to real time; this separation allows a clearer and more logical approach to the decision-making process in capacity planning (Billinton and Allan, 1985). We follow this separation in this work, and thus failure modes or constraints arising from the distribution and transmission networks are not considered.

The evolution of the system's adequacy is influenced by multiple factors. On the demand side, economic and technological change can alter power demand patterns through time; in the UK, a shift away from heavy industry, among other changes, has caused a downward trajectory in electricity demand since 2006, when the historic maximum was observed (Boßmann and Staffell, 2015; BEIS, 2016). On the other hand, most land transportation is projected to become electric by 2050 in all of the future scenarios considered by National Grid (NGESO, 2020b); moreover, it is also projected that by then an important share of heating will have switched from gas to electricity as well (NGESO, 2020a); these technological changes will increase electricity demand and likely alter consumption patterns considerably. Weather conditions can also impact demand levels; in the UK, the severity of cold spells during winter has a strong impact on observed peak demand (Richards and Ong, 2019), and it is reasonable to think that the sensitivity of demand to weather-related factors could be exacerbated even further in the future as heating becomes more reliant on electricity; moreover, climate change could also alter the frequency and severity of cold spells in unpredictable ways, making it harder to plan for future peak demand seasons, particularly considering the increasing adoption of weather-dependent renewable generation technologies (NGESO, 2020b). All of these exogenous sources of variation must be accounted for in order to produce credible demand scenarios with which to evaluate future capacity.

On the supply side, and for the purpose of adequacy models, it is useful to make a distinction between what is usually called *conventional* and *intermittent* power generating units. The former refers to units that are assumed to be dispatchable, in the sense that (in the absence of unexpected mechanical breakdown, which occurs probabilistically) their output level can be scheduled in advance as per economic considerations. This formulation is usually deemed appropriate for traditional fossil-fuel based units, such as gas turbines, coal, biomass or hydro

generation, assuming water reserves are available. Note that from an adequacy perspective, units are considered available as long as mechanical breakdown does not occur, even when the unit is offline, as it is assumed they can be brought online if the need arises.

Intermittent generating units are those whose output level cannot be scheduled in advance because of its dependence on exogenous factors such as meteorological conditions; wind turbines and solar panels are the most popular examples of this. As part of the decarbonisation efforts from international treaties such as the Paris Climate Accord, fossil fuel based generators are being decommissioned in GB and many other countries in favour of renewable energy sources like these. This transformation has widespread implications for power systems, from the structure of energy markets to public policy objectives (Morales González et al., 2014; Sensfuß et al., 2008), and the system's adequacy is not an exception: to preserve security of electricity supply throughout this transformation in the generation mix, in Great Britain a capacity market started auctions in the UK in 2014 and commenced operations in 2018. In it, utilities can bid to provide generation capacity to the grid operator between one and fifteen years into the future (Engie, 2016). Successful bidders receive payments in the year of delivery regardless of whether capacity is actually required for dispatch by the grid operator. The objective is to ensure there is sufficient investment in new capacity to meet ongoing reliability standards. Targets of capacity to procure for future years by the system operator are informed by the results of resource adequacy studies (NGESO, 2021c).

In addition to each system's installed capacity, interconnection to neighbouring countries and electricity trade across borders are becoming more common; besides the economic arguments, an advantage of interconnection is that it can enhance system reliability by potentially providing an additional source of energy when capacity reserves are low. In the context of the current energy transition, interconnection over wide geographical regions can offset some of the additional variability in available capacity introduced by intermittent generation at the national level (M. Grams et al., 2017). In 2020 there were three interconnectors in the UK to Ireland and France with a total capacity of 5 GW, and it is projected that by 2030 interconnection capacity will have at least a three-fold increase, with trading partners including Denmark, Norway, Ireland, France Belgium and the Netherlands (NGESO, 2020b). The trend toward more interconnected systems encompasses the European Union (EU) as well, where the 2013 TEN-E directive set a target interconnection capacity equivalent to 10% of electricity consumption for each member state (Council of European Union, 2013). The benefits

of interconnectors from the point of view of security of supply have long been recognised and studied (Billinton and Allan, 1985; Cook et al., 1963; Vassell and Tibberts, 1972); however, incorporating them into adequacy calculations can be a complex problem, and methodologies vary considerably across system operators: for instance, while in France and Belgium detailed simulations of neighbouring system at the level of unit commitment (i.e., scheduling the output level of individual generating units) are used to this end, some system operators in the US take available imports as fixed, and some other systems assume there are no interconnectors (National Grid, 2017b).

Resource adequacy assessments provide the tools to keep pace with technological and economic changes in the electricity system and reduce its carbon footprint without compromising consumers' access to electricity or risking damage to critical infrastructure; moreover, it is also aimed to prevent over-investment in capacity that would be reflected in unnecessarily high electricity bills for consumers. Research in resource adequacy methodology goes back to at least the 1930s (Benner, 1934; Smith, 1934), and practices have been developed on the basis of decades of experience by system operators. In this work, we look at some of the modelling practices in the Great Britain (GB) system, and develop methodological extensions with which we analyse the system's risk taking into account interconnection capacity and in the context of an increasing reliance on wind in the GB system.

1.2 Scope of this work

Many research questions are arising or becoming more prominent as the transformations mentioned above take place. This work focuses mainly on two challenges.

The first is devising principled probability models that account for extreme events relevant to power systems. This is important since most adequacy risks arise from unexpected occurrences of very high demand or very low available capacity; this problem is even more important when modelling multi-area systems, as sparsity of data in tail regions (this is, co-occurring extremes) is exponentially exhacerbated in multivariate settings due to the curse of dimensionality. In this regard, extreme value theory (EVT) offers mathematically principled results with which to model rare high-risk events and the statistical dependence between their occurrences across areas. Moreover, parametric extreme value models offer a smooth alternative to sparse data regimes from commonly used empirical models in security of supply; the latter may produce risk indices that are heavily

concentrated on a handful of past observations and thus suffer from high sampling variability.

The second challenge we focus on is performing an in-depth analysis of how relevant risk distributions change as the proportion of renewable generation grows to be the dominant factor in the generation mix. Great Britain, as well as other power systems around the world, are experience a quick shift towards a low-carbon generation mix, and it is crucial to understand how this transformation affects the risk profile of these systems in terms of adequacy, as this may have important consequences in decision-making for long-term system planning.

We go beyond usual univariate risk indices based on long-term averages (i.e., first moments) and take a wider look at how multiple relevant risk probability distributions change as the system moves to a high-wind regime. We also look at the current methodology for capacity procurement in Great Britain and incorporate a measure of risk aversion into the formulation (as it is likely decision-makers are not risk-neutral when it comes to preventing high-impact capacity shortfall events). With this methodological extension, we look at how the amount of installed wind capacity and the degree of risk aversion from decision-makers change capacity procurement decisions.

1.3 Contributions

The contributions from this work are methodological in nature, and are intended to be of practical use to practitioners and policy-makers; we do not develop further theoretical results in statistics or power system adequacy, and instead use real available data and existing statistical theory from various areas (such as the theory of extremes) and from the literature of power system adequacy to addresses the challenges outlined above in novel ways.

The outputs of this doctoral research project go beyond this thesis, and include:

- The publication of some of the contents of Chapter 3 in the proceedings of Probabilistic Methods Applied to Power Systems (PMAPS) 2020 (see Sanchez et al. (2020))

- The submission of a journal article with some of the contents of Chapter 4 to Sustainable Energy, Grids and Networks (SEGAN) (still in review at the time of this writing)

- Collaboration with National Grid on 2021 through the University's statistical consulting unit for the creation of a risk assessment study for the year of 2027.
- The creation of a public, high quality Python package with adequacy assessment functionality and implementations of some models from extreme value theory (see Sanchez (2022)).
- Collaboration with the Alan Turing Institute for the creation of a manual and a quickstart guide intended for R users for the package mentioned above.
- The submission of some of the contents of Chapter 5 to the proceedings of the IEEE Power and Energy Society General Meeting (PES GM) 2023 conference.

1.4 Thesis organisation

This thesis is organised as follows: in Chapter 2 we provide some historical background to this work and give a brief overview of the relevant literature and current modelling practices in power system adequacy that we use throughout this thesis. Chapter 3 outlines an analysis of the effects of different shortfall-sharing policies on post-interconnection risk in a two-area system using data from the Ireland-Great Britain system, and also analyses the impact of said policies and different choices of risk indices on the valuation of interconnection in the context of a growing reliance on intermittent generation. In Chapter 4, a modelling methodology based on extreme value theory is outlined for calculating post-interconnection risks in a two-area system, using the same data as in Chapter 3 to analyse the effects of statistical dependence across areas on post-interconnection risk and the value of interconnection. Finally, in Chapter 5, we move to single-area analysis and explore the evolution of the system's risk profile beyond long-term averages, focusing on aspects that are likely relevant for (possibly risk-averse) decision makers, again in the context of an increasing reliance of the Great British system on wind generation; we also incorporate a notion of risk aversion into the current capacity procurement methodology in Great Britain, and explore how parameters such as the degree of risk aversion and the estimated value of lost load may affect the optimal capacity to procure. Finally, Chapter 6 concludes this thesis by summarising our findings and the possible extensions to our work.

Chapter 2

Literature review and current practices

In the problem of adequacy assessment, the system's risk is usually calculated for a future scenario of what is connected to the system, and is usually measured using statistics of the corresponding stochastic process of capacity surplus/deficit (defining it as available supply minus demand across time) such as the expected duration of capacity shortfalls in a period of time. Most of the complexity in this problem come from devising appropriate statistical modelling methodology for both available supply and demand, as well as formulating credible future scenarios. This chapter offers a brief history of power system adequacy modelling and introduces some of the main risk indices used in the literature; after this, a summary of modern practices for modelling dispatchable generation, demand, wind generation and interconnection is outlined.

2.1 Historical context

Power systems have always been built placing emphasis on a reliable and affordable supply of electricity. In the interest of reliability, redundancies have been built into the system, and a question that arises naturally is then *how much* redundancy should be built. As additional redundancy increases costs, affordability and reliability are thus competing qualities that must be balanced. Power system reliability is a broad topic, and in this work we focus on adequacy, which refers to the existence of enough generation capacity to meet power demand. When it comes to generation capacity planning, the first techniques used to address this problem were deterministic in nature (Billinton and Allan, 1985). An example of these metrics is the *percentage reserve margin*, where installed capacity equals

the expected maximum demand plus some fixed percentage of it, or the *spinning reserve capacity*, where spinning capacity (this is, the aggregated capacity of all the generating units currently online) equals the expected load plus some fixed amount, usually equivalent to one or more of the largest generating units in the system; the latter could be used to prevent consumer disconnections in the event of the loss of the largest generator.

However, features of interest in power system adequacy such as generating unit failures and the evolution of electricity demand are probabilistic in nature. Hence, deterministic methods are not entirely appropriate in this context, and basing reliability standards on these criteria may produce lower -than-anticipated adequacy levels: in (Billinton and Allan, 1985) it is shown that systems that look identical under deterministic criteria like the ones described above can have markedly different capacity shortfall probabilities, depending on the particular composition of the generator fleet, i.e., the number and size of generating units, as well as their failure rates.

The need for probabilistic modelling in this area was recognised since at least the 1930s, but the lack of data and computing facilities at the time, as well as a lack of familiarity with probability theory by engineers and scientists prevented its widespread adoption (Billinton and Allan, 1984). An important group of papers in this regard was (Lyman, 1947; Loane and Watchorn, 1947; Seelye, 1949; Calabrese, 1947). They formalised the notion of probabilistic loss of load analysis by modelling the probability distribution of available generating capacity as the aggregated availability of individual generating units, which were modelled as independent, binary random variables (a similar model is described in Section 2.3). At first, calculations in these and other similar articles relied on assumptions such as identical failure rates and nameplate capacities across units; these simplifications were again in part due to the limited amount of data and computing power at the time.

A decade later, a subsequent group of articles (Baldwin et al., 1959b,a,c) proposed stochastic models for individual generating units with exponentially distributed failure and repair times. The referenced articles developed a comprehensive statistical analysis of generator failures, including uncertainty estimates around observed failure rates in historic data, under a range of assumptions. Moreover, they developed an efficient simulation methodology for relatively large fleets of generating units using these models, although still relying on some simplifying approximations.

In the following decade, Markov chains were also explored as models for individual generating units (Hall et al., 1968; Ringlee and Wood, 1969; Galloway

et al., 1969). In the context of the referenced articles, the main advantage of Markov chain models over those described in the previous paragraph is that they made possible, in a recursive manner, to construct a model for the aggregated available capacity from all of the generating units. Some results on the frequency and duration of aggregated capacity outage levels were then derived. Moreover, an analogous model for daily peak loads was also used to compute loss of load metrics for a hypothetical system with a small number of generators.

Models of available conventional capacity like the ones described above were combined with statistical models of the system's demand to compute quantities of interest related to capacity shortfalls (See section 2.2 for a review of risk indices). The formulation on which this is usually done is as follows: let X_t, D_t be stochastic processes representing available capacity and demand at time t. We call the difference the *capacity surplus* (*deficit*) and define it as

$$Z_t = X_t - D_t \tag{2.1}$$

Negative values indicate a capacity shortfall, which might produce involuntary customer disconnections, i.e., a loss of load. Unlike X_t, for which a probability models is fully specified using data on nameplate capacities and failure rates of generating units, like the ones described above, demand models are typically based on observed historic data, which can be forward-mapped in some way to the future scenario of interest (see Section 2.4 for more details); it is also usually assumed that X_t and D_t are statistically independent.

2.1.1 Incorporating wind generation into the problem

Initially, the formulation of the problem of adequacy assessment only accounted for conventional (i.e., dispatchable) generating units on the supply side. Currently however, renewable intermittent generation represents a significant portion of the generation mix in many power systems, and their share is only projected to grow. In the case of Great Britain, wind is the main source of intermittent generation (NGESO, 2020b). We can incorporate wind generation W_t into the formulation simply by updating (2.1) to

$$Z_t = X_t + W_t - D_t \tag{2.2}$$

However, care must be taken to model wind generation appropriately, as it does not fit within traditional models for conventional generating units. This is due to the fact that its output is determined by exogenous factors such as

meteorological conditions and cannot be scheduled in advance even in the absence of mechanical failure. Because of this, renewable generation is modelled as a time series which depends entirely on exogenous variables, just like how demand is modelled; this is in contrast to conventional generating units, for which a probability model is defined using failure rates and nameplate capacities (see Section 2.3).

There is evidence W_t and D_t might also exhibit a complex statistical dependence structure (Thornton et al., 2017). To avoid the complications of modelling this association directly, in this thesis we work with *net demand* $Y_t = D_t - W_t$, which we define as power demand minus wind generation. This is, in all of this work, we take net demand as the amount of capacity that has to be met by available conventional generation; this can be expressed algebraically as

$$Z_t = X_t - (D_t - W_t) = X_t - Y_t \tag{2.3}$$

This approach can be justified by the fact that in electricity markets renewable generation is often used up before conventional generating units are scheduled, hence the latter have to only satisfy net demand as defined above. This is because in general renewable generators have lower operational costs (i.e., ongoing costs throughout the plant's lifetime), since they do not need fuel to produce power and generally require less personnel, and hence usually sell their output at a lower price[1]. This is related to what is known as the *merit order* effect of renewable generation, which refers to the fact that the introduction of renewable generation into electricity markets tend to lower electricity prices for the reasons outlined above (Ballester and Furió, 2015; Antweiler and Muesgens, 2021).

2.2 Common adequacy indices

In adequacy studies, risk is usually expressed through statistics of the system's capacity surplus/deficit Z_t in the scenario of interest. One well-known risk indices is the loss of load probability (LOLP), which at time t is defined as

$$\text{LOLP}_t = \mathbb{P}(Z_t < 0) \tag{2.4}$$

This is valid only for a given point in time, however, whereas the objective is

[1]Capital costs of renewable projects and public policy around them can also affect how renewable generation is priced in energy markets. This is an active research topic that is beyond the scope of this work; see (Ferrari, 2021; Morales González et al., 2014) for a more in-depth discussion.

usually to evaluate the system's level of risk over a period of time, e.g. a future peak season. For this purpose, the period of interest is discretised at a given time resolution (more on this in Section 2.2.1) and the risk added up over the resulting partition. One of the most popular indices is the loss of load expectation (LOLE), which is defined as the expected count of discrete time intervals at which a loss of load event occurs (Billinton and Allan, 1985; Singh et al., 2018). If the period of interest consists of n segments, this can be written as

$$\text{LOLE} = \mathbb{E}\left[\sum_{t=1}^{n} \mathbb{I}(Z_t < 0) \cdot \Delta_t\right] = \sum_{t=1}^{n} \text{LOLP}_t \cdot \Delta_t \tag{2.5}$$

where subscripts denote the time index t, Δ_t is the time resolution (usually expressed in hours) and $\mathbb{I}(\cdot)$ is an indicator function, i.e., it equals 1 when the argument is true, and 0 otherwise.

A shortcoming of LOLE is that it does not make a distinction between shortfall sizes, placing the same weight on shortfalls that affect a few customers and those that affect millions. An alternative risk metric that addresses this problem is the expected energy unserved (EEU) defined as

$$\text{EEU} = \mathbb{E}\left[\sum_{t=1}^{n} \max\{0, -Z_t\} \cdot \Delta_t\right] \tag{2.6}$$

This is the expected aggregate shortfall size on the whole period. Note that both of these indices are based in long term averages (i.e., mathematical expectations). This in itself can be a drawback for some purposes, as they cannot convey changes in higher order moments that may be relevant for the question at hand. See Chapter 5.

2.2.1 Interpretation of different time resolutions for LOLE

Time resolution in adequacy calculations can differ across studies and system operators, which entails some subtle differences in the interpretation of the results (Stephen et al., 2022). In the case of LOLE, some systems, like in Great Britain and much of Europe, use an hourly resolution[2], while others, like the ones in North America, use a daily resolution by using daily peak demand data. LOLE is expressed in hours/year or days/year, depending on the time resolution used. Care must be taken in this regard, as results for different resolutions are not directly comparable, and assuming otherwise is a common source of confusion.

[2]Note that in the case of GB, available data is in half-hourly resolution, but LOLE and other results are expressed in hours.

To see this, let us imagine a system with a LOLE of 1 day/year, meaning the expected number of days with loss of load events in a given year is one. The loss of load event can be as short as one hour or as long as 24 hours, but should all happen within a single day. If we try to translate this directly to a LOLE of 24 hours/year under an hourly resolution, we would find that such a system would have on average 24 hourly loss of load events that would be likely spread on different days, with as many as 24 days of single-hour capacity shortfalls or as little as one day of (round the clock) shortfalls; the latter is the only scenario in which both daily indices and their naive translation to an hourly criteria would be equivalent.

2.2.2 Time-collapsed models

Despite what the time-indexed notation might indicate in (2.5) and (2.6), for expected value based indices like these we can further simplify the problem by obviating the need to model temporal dependence in Z_t. We can do this by defining a *non-sequential* or *time-collapsed* model Z' of the capacity surplus probability distribution in which we sample time steps uniformly at random from the period of interest. Samples from Z' are thus independent and identically distributed (i.i.d), and we have that (2.5) is equivalent to

$$\text{LOLE} = n \cdot \mathbb{P}(Z' < 0) \tag{2.7}$$

Analogously, (2.6) is equivalent to

$$\text{EEU} = n \cdot \mathbb{E}[\max\{0, -Z'\}] \tag{2.8}$$

Thus, finding a suitable time-collapsed or non-sequential model Z' for the capacity surplus is sufficient to calculate LOLE and EEU indices.

Nevertheless, it is noted in (National Grid, 2017b) that most of the surveyed system operators use sequential Monte Carlo based methodology (see Section 5.4 for an overview of sequential adequacy modelling) to estimate LOLE or EEU despite their time-collapsed nature, with only GB and PJM Interconnection in the east coast of the US using time-collapsed modelling. The use of sequential modelling for LOLE/EEU calculations can be advantageous to take into account additional system features that are not time-collapsed in nature, such as the use of energy storage (Sioshansi et al., 2022). This refers to the dispatch of capacity stored in devices such as batteries at times of low available capacity margins. However, this requires a preceding time window of enough available capacity to

charge storage devices. This sequential process cannot be adequately accounted for in a time-collapsed model, thus sequential modelling is necessary to compute the corresponding LOLE or EEU in this case.

2.2.3 Adequacy indices in practice

In 2017, an international survey of adequacy practices and methodologies was conducted by National Grid, and it was found that most of the surveyed systems used LOLE as their main risk indices for adequacy modelling (National Grid, 2017b). In the US, most systems use a reliability standard of 1 day in 10 years, also expressed as 0.1 days/year. Some systems in the US, however, use a literal translation of this standard into an hourly resolution resulting in 2.4 hours/year. This includes XCEL, Entergy and Portland General. On the other hand, systems in the north-west part of the US use a LOLP of 5% at a yearly resolution, meaning on average one year out of twenty having with loss of load events. However, this places no restriction on the number or frequency of loss of load events within such a year.

In most European systems, LOLE is also commonly used as the main reliability standard under an hourly resolution, as is the case in Great Britain (GB) (3 hours/year), Ireland (8 hours/year), France (3 hours/year) the Netherlands (4 hours/year) and other countries, some of which use LOLE in conjunction with other indices; for example, standards in Denmark use both LOLE and EEU (National Grid, 2017b). It must be noted that although these systems use the same time resolution in their standards, the survey concluded that due to methodological differences across systems on adequacy calculations, it was not possible to find a system whose reliability standards were directly comparable with those of GB.

In practice, it is almost always the case that reliability standards based on LOLE and EEU are used as *relative* adequacy indices by system operators: this is, they are used to provide a before-and-after comparison of the system's adequacy, instead of precise numerical predictions of loss of load occurrences (Billinton and Allan, 1985). For example, an increase in a system's year-to-year LOLE or EEU would indicate a proportional decrease in resource adequacy relative to the system's previous state, and vice versa. In reality, most systems experience a lower number of loss of load events than indicated by LOLE/EEU (National Grid, 2017b). Part of the reason for this is that even when a capacity shortfall is projected to occur, this does not mean customers will necessarily be disconnected, as system operators have a set of real-time remedial actions that can be

taken to mitigate it or avoid it altogether. In GB this includes temporary over-generation orders, voluntary load reduction by large customers, voltage reduction and emergency imports from other systems (Ofgem, 2013). Today these actions are not factored into LOLE/EEU calculations in Great Britain, but they were before 1985. At the time, the reliability standard was an annual LOLP of 4% (i.e., four years with loss of load events out of a hundred years on average) after emergency measures, with the corresponding pre-emergency measures LOLP being 19% (National Grid, 2017b). All in all, multiple factors can result in the real system experiencing a lower number of loss of load events than indicated by LOLE or EEU.

An advantage of relative indices is that they might be better suited for this problem, as absolute estimate of loss of load statistics would carry substantial uncertainty due to the large number of components and processes that would need to be taken into account. Doing so would likely compound numerical errors and modelling inaccuracies, making it difficult to obtain results with any reasonable certainty. Relative indices on the other hand help tracking the evolution of the system's adequacy using past performance as a baseline, and provide a more useful assessment in this sense.

2.2.4 Value of lost load (VOLL) and value based adequacy assessment

Research before 1970 mainly addressed the problem of modelling available generation and enabled the widespread calculation of LOLE and EEU, but did not address the question of what constitutes an optimal or even acceptable level of LOLE. Starting from the 1970s, considerable research has been done on what is now called value based adequacy assessment, which analyses adequacy levels in the context of economic considerations by placing a monetary value on loss of load events (Billinton et al., 1983). This is a difficult problem (not least because absolute loss of load indices are usually required), as there are multiple social (i.e., non-monetary) costs that have to be factored in, and even monetary costs might not be straightforward to quantify. Explored approaches include quantifying social costs from government agency figures and economic costs from local business activity indices (Kaufman and Daly, 1978); estimating the lost gross national product (GNP) as a result of energy not supplied (Shipley et al., 1972); using GNP/KWh as a conservative upper bound and wages/kWh as a more reasonable upper bound (Telson, 1975); measuring the cost as residential willingness to pay from price elasticities (Myers, 1978), among others.

Modern reviews on estimation methodology for VOLL can be found in (Gorman, 2022; Schröder and Kuckshinrichs, 2015). Importantly, (Gorman, 2022) also discusses three of the main methodologies that have emerged throughout the last decades to quantify the VOLL, namely proxy, revealed preference and survey methods. Briefly, in order to estimate VOLL, proxy methods use macroeconomic indicators as a reference, revealed preference studies observe the behaviour of electricity consumers in the markets, while survey methods involve interviewing individuals or businesses to ask them for their specific VOLL or related measurements.

It is noted in (Gorman, 2022) that survey methods have historically been the most popular ones to estimate VOLL, as it is in fact done currently by Ofgem in GB (Ofgem and DECC, 2013). The preference of survey methods over the others have to do with many factors, among them known drawbacks and difficulties inherent to the other methodologies, as pointed out in (Gorman, 2022).

Within survey methods, there are a variety of approaches to enquire for the customer's VOLL, for instance, asking for their willingness to pay (WTP), this is, how much the customer is willing to pay in order to avoid a power outage with certain characteristics; another possibility is asking for their the willingness to accept (WTA), this is, the amount of money the customer is willing to accept to incur in an outage with certain characteristics. In theory, both should be equal (Hartman et al., 1991), but in practice both can end up being be very different, evidence of well known issues with these types of survey studies (Hausman, 2012).

A more robust result can be devised using conjoint analysis (Steiner and Meißner, 2018), in which customers are instead asked to order their preference between a set of alternative hypothetical outage scenarios. This approach is based on evidence that shows that indirect measurements of consumer preferences are better than direct measurements at predicting and understanding consumer behaviour (McFadden, 1986). Conjoint analysis has been used in the Netherlands to determine VOLL (Ecorys, 2022), and has recently been used by Ofgem in Great Britain to understand consurmers' makret behaviour (Ofgem, 2018).

Recent studies have been conducted in different countries to produce up to date estimates of the cost of energy not supplied similar to those discussed above. In (National Grid, 2017b) it is mentioned for instance that in France a study from 2012 calculates a cost of 26,000 euro/MWh, while customer surveys in Australia calculated a value of 33,460 AUD/MWh. Meanwhile for the UK, a survey conducted in 2013 calculated the cost at around 17,000 £/MWh.

It was noted in (National Grid, 2017b) that the use of cost-benefit analyses was not the norm among system operators (with GB being an exception to this

back then; see Section 5.3 for the current cost-benefit formulation for capacity procurement), however, in 2018 the European Commission formally mandated the development of reliability standards relying on cost-benefit analyses (EU, 2018). To the best of our knowledge this has not been implemented in any EU member yet, and some of the possible reasons for this are the high level of uncertainty involved in estimating relevant parameters, as discussed above, and the difficulty of obtaining absolute loss of load statistics, which are often required in cost-benefit analysis calculations. Instead, adequacy standards have historically been chosen by expert judgement, which could be seen as procuring an acceptable level of reliability at an acceptable cost, based on the system's historic performance.

2.3 Modelling conventional generating units

2.3.1 Time-collapsed models

The simplest and most common model for available conventional (i.e., dispatchable) generation in reliability studies assumes generating units can be either fully available or fully unavailable (i.e., no de-rated states exist) with some probability; furthermore, under a time-collapsed view (see Section 2.2.2) sequential dependence between their states can be ignored, which results in their availabilities being modelled as binary random variables (Billinton and Allan, 1985; Singh et al., 2018); this model has been used, for instance, in (Tindemans et al., 2019; Wilson and Zachary, 2019; Sanchez et al., 2020); in this Section we give a brief description of it.

In order to define the model above, data on the units' outage rates are needed. The outage rate of a unit or group of units is defined as the number of outage occurrences per unit of service time (IEEE, 2019a). For instance, if the service time is one year, using an hourly resolution this would be the number of hours of unavailability divided by the total number of hours in a year. Depending on the purpose of the study, different types of outages can be considered, e.g. only forced outages which occur due to unforeseen mechanical failure, or including scheduled outages due to routine maintenance; these rates can be calculated for an individual unit but it is common to do so for a class of units, e.g. nuclear generators.

A generator with an outage rate of π and nameplate capacity $c \in \mathbb{R}$ is modelled as a random variable G such that

$$\mathbb{P}(G = x) = \begin{cases} \pi, & x = 0 \\ 1 - \pi, & x = c \end{cases} \tag{2.9}$$

The available capacity of a fleet of conventional generating units would consequently be given by the aggregate available capacity of all of its units. For a fleet with m generating units this can be written as

$$X = \sum_{i=1}^{m} G_i \tag{2.10}$$

where G_i denotes the availability of the i-th generating unit with a maximum capacity and failure probability of c_i, p_i, respectively. Note that there are no time indices, as we are assuming a time-collapsed view of conventional generation availability.

In general, the probability distribution of X depends not only on nameplate capacities and outage rates of individual units, but also on the statistical dependence across units. It is common in adequacy studies to assume statistical independence between generating units, as we do in this work. However, other dependence structures have been explored, and there is some evidence forced generator outages might be correlated in some circumstances. For example, in (Murphy et al., 2018) the assumption of independence between generators is analysed and evidence of statistical dependence is found in all regions. Temperature was determined to be one of the reasons behind this dependence, which could also be associated with natural disasters such as draughts and cold spells that affect wide geographical areas. A relevant example of this effect are the events in Texas on February 2021 in which multiple generators failed due to lack of fuel which was caused by frozen gas pipelines. Note that outages like these are weather-driven, and as such they could be taken as independent *conditional on* the weather for modelling purposes. This case is contemplated in IEEE standards for reporting outages of electrical facilities (IEEE, 2019b). Nevertheless, the authors note that even when excluding natural disasters from the data, evidence of dependence remains. Another example of this analysis is (Murphy et al., 2020), where a model for temperature-dependent correlated outages is used to assess increases in capacity shortfall risks due to a warmer future climate in some regions of the U.S.

For the model of conventional generation X in (2.10), and assuming statistical independence between generating units, the probability distribution of X can be computed from (2.9) alone by convolving the discrete probability distributions of

all generating units. To do this, nameplate capacities are usually rounded to the nearest integer for convenience. This is not an issue in practice, as nameplate capacities are usually in the range of tens to hundreds of MW, and so the error induced by rounding is negligible. Algorithms to numerically compute this probability distribution are outlined in (Billinton and Allan, 1985). Note that the convolution of integer probability distributions can also be efficiently done using a fast Fourier transform algorithm (Elliott, 1987).

2.3.2 Time-sequential models

Many time-sequential models have been proposed too for the availability of conventional generation capacity; in (Baldwin et al., 1959b,a,c), generating units are assumed to have exponentially distributed times to repair and times to failure. Another common method assumes the availability of generating units follows a Markov chain process (Billinton and Allan, 1985; Singh et al., 2018). The latter model have been used in (Edwards et al., 2017; Sheehy et al., 2016), and we also use it in Chapter 5 to calculate the output of sequential adequacy models. Note that in all of these, statistical independence between generating units is assumed. In this Section, we describe Markov chain models for generating units in more detail.

The simplest sequential Markov chain model in this regard is again one without de-rated states in which generators are fully available or not at all. This model can also be characterised as a two-state birth-death process as we show below; in what follows, rate units are omitted but assumed to be identical across parameters (e.g., hourly rates). In addition to the unit's outage rate, we need to know the unit's mean time to repair (MTTR), usually expressed in hours. To illustrate this, let G be a generator with an hourly outage rate of π and an MTTR of $\tau \geq 1$ (since the shortest time a Markov chain model can remain in a given state is a single time step). Then, the probability of transitioning to an availability state whenever the unit is unavailable is $\mu = \tau^{-1}$. Let λ be the probability of transitioning to an unavailability state whenever the unit is available. We can derive λ by noting that (2.9) gives us the chain's stationary probability distribution, and so if A is the chain's transition matrix, we must have that

$$A^T \boldsymbol{\pi} = \begin{pmatrix} 1-\mu & \lambda \\ \mu & 1-\lambda \end{pmatrix} \begin{pmatrix} \pi \\ 1-\pi \end{pmatrix} = \boldsymbol{\pi} \tag{2.11}$$

where the chain's states are $(0, c)^T$; this yields

$$\lambda = \mu \frac{\pi}{(1-\pi)} \tag{2.12}$$

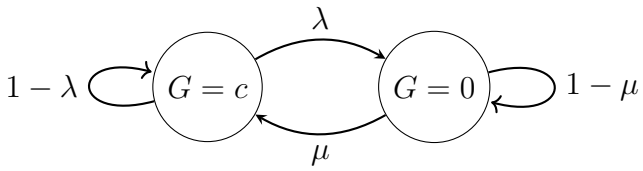


Figure 2.1: Birth-death process for a conventional generating unit. Edges are labelled by their corresponding transition probabilities.

The aggregate available conventional generation X_t would then be a sum of Markov chains, analogous to (2.10); moreover, statistical independence across generating units is also assumed. For time-sequential risk indices, Monte Carlo simulation is the only viable alternative for calculating model outputs.

The inclusion of de-rated states to both of the models outlined in this and the previous Section is straightforward from a computational and mathematical point of view, as this would just entail using different probability mass functions in (2.9) or larger transition matrices in (2.11). Nevertheless, sometimes system operators do not have access to the necessary data to specify more detailed models of generating units. Available data varies considerably across system operators. For example, power systems in North America use the Generating Availability Data System (GADS) as a centralised monitoring systems where comprehensive availability data is compulsorily reported for all units above 20 MW of capacity (NERC, 2022). The granularity of this data allows in-depth statistical analysis of the outage probability distributions both for individual units or unit types and for entire generator fleets, as has been done in (Murphy et al., 2018). On the other hand, in systems like GB there are no analogous standards and existing availability data is more limited (National Grid, 2017b).

There is not much discussion in the literature on how appropriate Markov chains are when modelling generating units. It is possible that two-state Markov models are too simple to be an accurate model for some types of units, e.g. those with multiple generating components that work in parallel and that could in principle fail independently. However, the inclusion of de-rated states could address this. Even though there is no evidence to our knowledge that mechanical failures are Markovian in nature, under the usual framework of characterising availability by mean times to repair or failure (or more generally state exit times, if more availability states exist in the Markov chain model) Markov chains with appropriate de-rated states would in principle be a valid model for this problem.

2.4 Modelling electricity demand

Adequacy calculations require the elicitation of credible future demand scenarios; demand data in the historic record deemed to be representative of current and near-future demand trends is generally used as a basis for this (National Grid, 2017b). The data are usually transformed in some way to represent a future scenario of demand. There are many ways of doing this, the simplest being rescaling by a suitable constant factor; this method is used at present by National Grid (National Grid, 2017b), and will be explained in detail later on. However, it has been pointed out this methodology does not account for things such as technological change and other effects that could alter the probability distribution of demand beyond a change in scale (Boßmann and Staffell, 2015). More complex modelling approaches have been explored in the literature for modelling future demand, for example, extrapolating trends in the historical evolution of load shape (Basaran Filik et al., 2009). Modelling future demand of particular technological transformations such as the adoption of electric vehicles or heat pumps (Koreneff et al., 2009); partitioning demand into different segments such as industrial and residential users, which are modeled individually and then added up to form an aggregated demand forecast (A. Pina, 2011), among others. We note that with the possible exception of (Fan and Hyndman, 2012), most of the recent models for demand in future years have been developed for purposes other than adequacy modelling, e.g. studying the sensitivity of future demand to different aspect of technological change, as done in (Boßmann and Staffell, 2015).

2.4.1 Using average cold spell factors to rescale demand in GB

In Great Britain, constant rescaling factors are used to forward-map historic demand to the future scenario of interest, which is usually a future peak demand season. These are called average cold spell (ACS) peak factors, and are calculated for each historic peak demand season individually, which in GB occur during winter. Generally speaking, they are meant to be a reflection of the effect of weather variability on the demand level conditioned to the corresponding peak season. More specifically, they are estimates of the median of the peak demand season's maximum half-hourly demand across a range of weather conditions. In the case of forward-looking adequacy calculations, if a future peak season under analysis is projected to have an ACS peak of α, then a historic peak season with an estimated ACS peak of β is forward-mapped to the future year's conditions by

rescaling observed demand values by α/β. Once historic peak seasons have been standardised in this way, the data is deemed to be representative for the scenario of interest.

The estimation methodology for ACS peak factors is detailed in (Richards and Ong, 2019). Briefly, a linear model for total demand conditioned to the corresponding peak season is defined with the following structure

$$\begin{aligned}\text{total demand} =& \text{base demand} + \text{temporal effect} + \\ & \text{weather effect} - \text{unmetered generation} + \\ & \text{random variation}\end{aligned} \tag{2.13}$$

Here, the left hand side represents demand metered at the transmission level, i.e., demand that has to be met by the generation subsystem and moved through the transmission system to the customers. On the right hand side we have base demand, which refers to the underlying level of non-varying (i.e., constant) demand independent of other factors during the corresponding peak season; temporal effects, which account for variations due to day of year and day of week; weather effects measures dependence on weather conditions, particularly temperature features; unmetered generation is generation embedded in distribution networks (more on this below); finally, random variations account for residual variability.

We note that unmetered generation is subtracted on the right hand side; unmetered generation refers to generation embedded into distribution networks (e.g. solar panels installed on a customer's rooftop) that can satisfy local demand without making use of the transmission network. For this reason, this needs to be subtracted from total demand in (2.13) to estimate demand metered at the transmission level.

The model above is used in a Monte Carlo estimation process to calculate ACS peak factors. To do this, historic weather data are block-resampled in weekly blocks to assemble a peak season's worth of synthetic weather conditions. These are used as inputs to calculate weather effects and unmetered generation in (2.13), and this demand model is then used to produce a demand trace for the synthetic peak season. This is repeated 20,000 times, from which the median of the maximum half-hourly demand is computed.

For each historic peak season, both a forecast and outturn ACS peak values are calculated before the season starts and after it ends, respectively. Forecast ACS peak values are published in a yearly Winter Outlook report by National

Grid right before the season's start. For the outturn ACS peak, features such as the base demand level and temporal effects are updated using observed data, and the estimate is recalculated. Note that in contrast to years in the historic record, estimating the ACS peak for a peak season multiple years into the future which is usually the case for the scenario of interest might require the use of market intelligence and expert judgement.

Recently, linear models that account for season-specific weather and temperature features have been proposed as a more expressive alternative to the ACS peak methodology. In (Wheatcroft, E. and Dent, C. and Wilson, A., 2022), linear models with weather and temperature features are fitted to each historic year and used to project the weather conditions of one year onto demand from a different year. The linearity of the models allows doing this in a straightforward way by simple addition and subtraction of estimated weather parameters for different years, while allowing for a richer set of credible scenarios compared to a constant rescaling factor. The authors point out that a demand model like (2.13) could be plugged into this methodology to achieve more expressive demand forecasts in forward-looking calculations without the need to block-resample historic winters, which could be problematic (e.g., random block resampling might implicitly assume that conditions at different points in winter are interchangeable).

Eirgrid, the Irish system operator, follows a similar methodology as the one outlined above for modelling future demand scenarios into their reliability studies, while in other countries methodological details generally differ and are specific to each system operator (National Grid, 2017b). For example, the cited survey mentions that PJM in the US scales historical loads by a random factor in a range of values determined by expert judgement.

2.5 Modelling wind generation

In the last twenty years, wind generation has grown to encompass an important proportion of the generation mix in the UK and other European countries. However, unlike conventional generating units, the output of wind farms cannot be scheduled in advance and instead depends on weather conditions. This creates the need for accurate historic wind generation data for adequacy calculations. Nevertheless, generation data from wind farms are typically not public.

Partly because of this, it is a common practice in adequacy studies to use public weather reanalysis models such as NASA's MERRA (Rienecker et al., 2011) or MERRA2 (Gelaro et al., 2017) to reconstruct historic wind generation traces

for the wind fleets in the scenario of interest. Moreover, this practice also provide a consistent historic record to explore hypothetical wind scenarios. Reanalysis models combine satellite and ground-based historic weather observation data with short-range weather forecast models to provide an approximation to the planet's weather conditions at an evenly spaced grid on the Earth's surface. This is done across time at short time intervals (e.g. hourly), and in the case of MERRA2, the grid resolution is roughly 60 km. Wind speed data on this grid can be interpolated to wind farm locations, and the results can be combined with data on the type and number of turbines at the wind farm to estimate the farm's hourly capacity factors, e.g. its output expressed as a proportion of installed capacity. These estimates are subsequently aggregated at a national level in the form of national capacity factors, which can then be rescaled to actual wind generation output using data on the amount of total installed wind capacity under the scenario of interest.

However, reanalysis data needs careful validation and calibration, as it has been found that raw reanalysis data incur in significant spatial bias when used to this end (Stickler and Brönnimann, 2011; Decker et al., 2012), due in part to the relatively coarse grid over which measurements are imputed, which could overlook blocking geographical features affecting wind farms, among other issues. In (Staffell and Pfenninger, 2016) it is found that a consequence of this is an overestimation of wind capacity factors in northern Europe by roughly 30% while underestimating those of southern Europe; the cited paper develops a bias correction methodology which uses an affine transformation of reanalysis data's wind speeds (i.e., a change in the data's location and scale) so that simulated historical wind capacity factors for on-shore and off-shore fleets match long-running observed historical averages. The results show high goodness of fit, achieving close to 95% correlation with observed hourly wind speed validation data in northern Europeean countries, including the UK; the resulting corrected wind speed data is openly available; see (Staffell and Pfenninger, 2016).

2.6 Modelling multi-area systems

In this work, we call multi-area systems those that share an interconnection through which power can be exchanged and which are managed by independent system operators. Many power systems around the world are part of broader multi-area systems as defined above, such as those in North America (except Texas' ERCOT) and most of Europe. Being part of a multi-area system gener-

ally improves adequacy, as systems experiencing tight margins can be assisted by other areas if spare capacity is available. This is in addition to other economic benefits from the resulting market coupling; see (Newbery et al., 2019) for a more detailed analysis of their economic impact.

Multi-area adequacy models are also called multi-node models, as they are usually represented as an undirected network where each node represents a system and edges represent tie lines, or interconnectors, between them. In a multi-area system, the risk in any area depends now on the resource-sharing policy between system operators, the load and available generation at other areas (since this determines whether any assistance can be provided), and interconnection capacities.

Resource-sharing policies dictate how interconnection would work when one or both systems experience a capacity shortfall. Two of the most common ones in the literature are the *veto* or *no load loss sharing* (NLLS) policy and the *share* or *load loss sharing* (LLS) policy (Singh et al., 2018; Tindemans et al., 2019). The former means power can flow to other areas only after domestic load is met; under a share policy on the other hand, loss of load is shared across areas to the extent interconnection permits. The latter would be, in principle, compatible with non-discrimination regulations that the EU aim to implement in its energy network (Rumpf and Bjørnebye, 2019). See Section 3.3 and Figure 3.2 for the version of these policies we use in this work and their concrete mathematical specification.

Multi-area models have been studied since at least the 1960s: for instance, in (Cook et al., 1963), LOLP calculations for a 2-area system are outlined, assuming a veto policy (i.e., only spare capacity after meeting local load can be exported), and are used to evaluate the reduction in reserve requirements in a small 2-area system. In (Vassell and Tibberts, 1972) a similar approach is used to study extensions to three and more systems, with and without loops. However, simplifying assumptions about the order of interaction between systems were used, e.g., a fixed order of priority when exporting power to systems that need it. This can help simplifying the analysis if complex topologies are involved; moreover, pairs of areas are merged by convolving their capacity probability distributions, implicitly assuming statistical independence. In (Pang and Wood, 1975), models for multi-area system with an arbitrary topology are also explored, this time without simplifying assumptions about the model's topology, using a min-cut algorithm to determine actual power flows between systems instead; however, statistical independence is again assumed between individual systems.

One of the easiest and most popular approaches for incorporating neighbour-

ing areas into a system's adequacy calculations is treating them as two-state generating units (see Section 2.3) whose availability probability distribution is then convolved with the local capacity surplus/deficit probability distribution to compute risk indices (thus, once again assuming statistical independence between systems). This has been called the equivalent unit approach or equivalent assistance approach (Billinton and Allan, 1985; Singh et al., 2018). More formally, let Z be the capacity surplus in the local area. Then under the equivalent unit approach a random (usually binary) variable G representing the capacity at the other side of interconnection is added to Z, hence the post-interconnection capacity surplus of the local system is given by $\mathcal{Z} = Z + G$, which can be computed through the convolution of the associated probability mass functions. This approach can use a two-state or multi-state generating unit probability distribution for G. Note that under such a model, modelling particular shortfall-sharing policies is problematic, as by construction it is assumed available imports are independent from local capacity surpluses/deficits. If there are multiple neighbouring systems, all of them would be modelled analogously. This approach is computationally cheap. However, as was said above, it implicitly assumes both systems are statistically independent. This assumption can be problematic, for instance if the system have similar daily demand patterns (as is usually the case in interconnected systems, since areas are geographically close), as the resulting models likely overestimate the utility of interconnection in adequacy calculations.

A more detailed approach is using the full multivariate capacity surplus/deficit probability distributions for risk calculations, where each dimensional component correspond to one individual system. This is also sometimes called the probability array or contingency enumeration approach (Billinton and Allan, 1985; Singh et al., 2018). For a k-area system this would require computing the multivariate capacity surplus probability distribution $\boldsymbol{Z} = (Z_1, ..., Z_k)$. In numerical calculations, each univariate capacity surplus is usually modelled as a discrete random variable with an integer support (which can be the result of convolving the probability distributions of available conventional capacity with demand in (2.1) or net demand in (2.3)), and this can be in the thousands to tens of thousands in length (as they are commonly expressed in MW). Thus, exact calculation of post-interconnection risk indices would entail sweeping the corresponding k-dimensional grid. For these reasons, exact computations using this model can quickly become impractical as more neighbouring systems are added, and can be computationally expensive even for two-area systems; binning the support of capacity surpluses/deficits into coarser resolutions, as done in (Tindemans et al., 2019), might help keeping this model computationally feasible. See Chapter 3

and Appendix B for implementation details of this approach in a two-area systems at full resolution for the versions of veto and share policies explored in said Chapter.

Other approaches have been proposed to find a compromise between the model's level of detail and computational efficiency. For instance, in (Lago-Gonzalez and Singh, 1990), conventional generation states are aggregated into disjoint sets, and multivariate demand observations are merged into clusters that are intended to represent different types of load scenarios. Monte Carlo simulation can also be used on top of these decomposition methods or in conjunction with them to find relevant sets of system states (Clancy et al., 1983; Singh and Deng, 1991; Singh and Mitra, 1997); more recently, Monte Carlo methods have also been used directly on top of full multivariate models for capacity surpluses/deficits (Singh et al., 2018).

2.6.1 Capacity value of interconnection

Methods have also been proposed to estimate capacity values metrics for interconnectors. The objective of capacity value metrics is to provide a common basis to compare the contributions to system adequacy from assets as different as conventional generating units, battery storage, renewable generating units and in this case interconnectors (Tindemans et al., 2019). Two common capacity value metrics are equivalent firm capacity (EFC) and equivalent load carrying capacity (ELCC), both of which are computed relative to a user-chosen risk metric, e.g. LOLE or EEU (see Section 2.2) and will be briefly explained below; see (Zachary and Dent, 2012) for a more in-depth discussion of capacity value metrics.

The EFC is defined as the amount of firm capacity (i.e., constant available capacity) to which the asset in question (say, a conventional generating unit, a wind farm or an energy storage device) is equivalent in terms of the chosen risk metric. In other words, installing said asset into the system or adding its EFC on the supply side are equivalent alternatives with respect to the (user-chosen) system's risk metric.

ELCC on the other hand expresses an equivalence in terms of additional firm demand rather than supply. This is, adding an asset (say, battery storage) to a system's supply and at the same time its corresponding ELCC as additional firm load (i.e., constant load) would leave the chosen risk metric unchanged. Although conceptually both EFC and ELCC could be seen as similar in some sense (as both represent firm added capacity but with opposite signs), this is not mathematically true due to the non-linearities induced by risk indices. However,

it has been proven that when the new asset's capacity is small relative to what its already installed in the system, EFC and ELCC essentially coincide (Zachary and Dent, 2012); however, this may not hold in the case of interconnectors as their capacity is usually in the order of GWs.

To see how both metrics are calculated in the case of interconnectors from the point of view of the local system, let $r(\cdot)$ be the user-chosen risk metric as a function of the system's surplus, e.g. LOLE or EEU in (2.5) or (2.6) respectively, and let $Z, \mathcal{Z}$ be the system's pre and post interconnection capacity surplus/deficit. Then interconnection EFC and ELCC are such that

$$r(Z + \text{EFC}) = r(\mathcal{Z}),\ \text{EFC} > 0 \tag{2.14}$$

$$r(\mathcal{Z} - \text{ELCC}) = r(Z),\ \text{ELCC} > 0 \tag{2.15}$$

We emphasise that the Equations above value interconnection from the point of view of the local system only, and assume the neighbouring system will not adapt to the presence of interconnection. However, computing simultaneous interconnection valuations when both systems take advantage of it (e.g. change their conventional generation fleet) is more complex, as changes in one system will change interconnection value for the other system. In this sense, a methodology to compute the Pareto frontier of interconnection valuations based on ELCC was developed in (Tindemans et al., 2019). Furthermore, (Greenwood et al., 2022) develops the concept of minimum EFC to value interconnection; this is the smallest amount of firm generation that has to be installed across both areas (possibly unevenly) to replace interconnection in terms of adequacy risks.

Note that EFC and ELCC are used inside pre and post-interconnection risk indices respectively; note also that ELCC is subtracted as it represents additional load to the system. In (Tindemans et al., 2019), methodology is outlined to compute interconnection's ELCC using data from Great Britain and France. Computing interconnection's capacity value metrics is important to determine to which extent they can participate in capacity auctions, for instance. Finally, the calculations of quantities involved in (2.14) and (2.15) can be performed through any of the techniques discussed earlier in this Section.

2.6.2 Other benefits of interconnection in system adequacy

In the current context of decarbonisation, interconnection can dampen variability arising from a growing reliance on intermittent energy sources. In (M. Grams et al., 2017), strategies for deployment of wind farms throughout Europe are explored, aiming to reduce variability in continental-level wind generation. Strategies like this one would make heavy use of interconnectors to transfer large amounts of power between northern and southern Europe, which are negatively correlated in terms of wind generation.

Interconnection would also allow, in principle, a more efficient allocation of resources when procuring new capacity across multiple areas. In (Hagspiel et al., 2018), they show how a system representative of the European network could reach target adequacy standards while procuring 6% less capacity by using a coordinated approach compared to an uncoordinated one. The authors note that changes in system-wise adequacy targets can have a considerable impact on the results, which suggest standardisation of adequacy targets and methodology across areas could be beneficial in this coordinated scheme. It is unclear, however, how closely the numerical results would be reproduced in the real European network, as some of the assumptions in the study might not be representative of the system. For instance, the authors assume demand and wind generation data are essentially independent, using different year combinations from both data sets, which might mischaracterise statistical dependence between wind and demand; moreover, available conventional generation seem to be taken as constant in all areas rather than as probability distributions. However, the results do offer motivation for further study on the benefits of adequacy coordination.

2.6.3 Multi-area calculations in practice

In practice, methodology to incorporate interconnection into adequacy calculations varies across systems, as described in (National Grid, 2017b). In Great Britain and Finland, BID3, a market dispatch model with specification for multiple European power systems, is used to this end, while Belgium and France use Antares, an open source software similar in functionality to BID3. The equivalent unit approach, described in this section, is used by operators in Spain, Denmark and Ireland, while others do not include interconnection in the calculations, like in the Netherlands. We note that using market dispatch models for this purpose could be problematic, as these assume normal operating conditions, while

adequacy calculations are mainly driven by rare, extreme events. In Chapter 4, we propose an alternative methodology for 2-area systems based on extreme value theory, a branch of statistics that offers principled models for characterising extreme events, using data from the Ireland-Great Britain system.

Chapter 3

Quantifying The Reliability Contribution of Interconnectors in the Britain - Ireland Power System Using a Hindcast Approach

A reliable electricity supply is a key consideration for energy system planners. At present, the value of support from other systems is of particular interest in Great Britain (GB), including from Ireland's (IRL) Single Electricity Market, to which at the time of this writing there is 1 GW of interconnection capacity. This chapter presents a study of how interconnection influences risk levels in the GB and IRL systems, based on the standard Loss of Load Expectation (LOLE) and Expected Energy Unserved (EEU) indices, and a *hindcast* approach for demand and wind generation. Specific areas of investigation include the effect of different resource sharing policies on adequacy risk levels under a range of assumptions regarding renewable penetration and interconnection size, as well as the dependence of the value of interconnection on the wind capacities in the two systems.

3.1 Introduction

Along with sustainability and affordability, security of supply is one of the three pillars of energy policy. In Great Britain as in many other power systems, there is thus considerable interest in the outputs of security of supply risk calculations, and development of methodology for this appropriate to any new circumstances –

such calculations may be for the general purpose of understanding the risk profile of the system, or they may be performed as part of the operation of a capacity market (NGESO, 2019).

At present there is increasing interest in the contribution of interconnection between systems to security of supply – this is due to the possibility of available renewable capacity being less variable when aggregated over wider geographical areas, and of political drive for market integration (European Comission, 2016), in addition to the economies of scale available from sharing planning reserve over wider areas which have long been recognised.

Quantifying the contribution of interconnectors is a long-standing area of study, and has become a standard topic in the power system reliability literature (see Section 2.6 for an overview). A recent development in this topic has been the study of how capacity value metrics may be used to value interconnection between systems (see Section 2.6.1), including consequences of different policies for sharing resources (for instance comparing a policy where each area looks after itself first, to one where at times of shortfall resources are shared between systems). In this work, we investigate contribution of the interconnection between Ireland (IRL) and GB to both areas. We work in a similar overall picture to (Tindemans et al., 2019), although with a somewhat different focus. The cited paper explores the valuation of interconnection for different resource sharing policies, characterising the set of Pareto optimal ELCCs for both interconnected areas. We analyse the valuation of interconnection as installed wind capacity grows, and look at how this valuation depends also on the combination of risk indices chosen to measure it and the resource sharing policy, taking the definition of a share and veto policies as described in (Tindemans et al., 2019). Furthermore, we compare the evolution of risk indices at both areas for increasing interconnection capacities under these policies. We model both systems jointly (see Section 2.6), and consider all demand and available capacity states at both; while this approach is also used in (Tindemans et al., 2019), binning the resulting bivariate model into a coarser resolution was necessary to control computational complexity. We derive exact formulas for EEU and LOLE for both of the considered policies in terms of univariate integrals, which allows for calculations at full resolution.

This chapter is organised as follows. First, Section 3.2 presents the two area risk model, and Section 3.3 describes the approach used to calculate risk indices. Section 3.4 describes the data used, and Section 3.5 presents numerical results. Finally Section 3.6 discusses issues arising from data limitations and possible future work to mitigate this, and Section 3.7 presents conclusions.

3.2 Formulation

In the following sections we denote random variables as uppercase, specific instances as lowercase, and vectors as bold. We define net demand as demand minus renewable generation and capacity reserves as available supply minus demand; we call these reserves capacity surplus/deficit. We assume the future period under study is divided into n time intervals (e.g. hourly), and then for a given time t the system's capacity surplus/deficit Z_t is given by the difference between available capacity and demand, i.e., $Z_t = X_t + W_t - D_t$, where X_t, W_t, D_t represent available conventional generation, renewable generation and demand, respectively. We further assume that conventional generation X_t is independent of all else, and that each generator can be either fully available or under forced outage (see Section 2.3). Note that in the case of renewable generation, we do not model plant outages. Using this framework we can calculate the standard risk indices LOLE $= \sum_{t=1}^{n} \mathbb{P}(Z_t < 0) \cdot \Delta_t$, or loss of load expectation, the expected number of discrete time units in which a shortfall occurs during the period of interest, and EEU $= \sum_{t=1}^{n} \max(-Z_t, 0) \cdot \Delta_t$ or expected energy unserved, in the period of interest. Without loss of generality, the time resolution Δ_t in these calculations is taken as 1h.

Typically one has available for statistical estimation historic data from multiple years of demand and available renewables (d_τ, w_τ), where τ indexes historic times, and these data combine historic weather and time of day/week/year with a scenario of what is connected to the power system in the future season or year under study. We assume that a distribution of available conventional capacity for that scenario is also available. This is usually modelled as in Section 2.3. The most common estimation approach for estimating the above indices is then 'hindcast', which considers what the risk level in the future scenario would be given a repeat of the historic weather (and thus also demand patterns). We then have an estimate of the LOLE not conditional on any particular weather conditions LOLE $= (1/s) \cdot \sum_\tau \mathbb{P}(X_\tau < d_\tau - w_\tau) \cdot \Delta_t$, where the sum is over the entire historic dataset and s is the number of years in it; an analogous expression may be written down for hindcast EEU. It is also possible to calculate a hindcast LOLE or EEU conditional on repeat of weather from a single specific historic year (Wilson et al., 2018).

3.3 Determining 2-area system risk indices

In order to do a proper risk evaluation in a multi-area system, it is necessary to take into account the capacity available to be imported from other areas at times of shortfalls. How much is available depends on what the capacity surplus/deficit in each area would be in the absence of interconnection, the policies in place for sharing power at times of shortfall in one system, and the interconnector capacity. From a security of supply perspective, we are interested only in the **available** power to support neighbouring systems, not what will flow in day-to-day trading when no system is under stress. Finally, we assume that full interconnector capacity is available throughout the peak season.

In a 2 area system, the available capacities and demands are now vectors $\mathbf{X}_t, \mathbf{W}_t, \mathbf{D}_t$ of length two. It is convenient first of all to look at the risk calculation for a single fixed demand $\mathbf{d} = (d_1, d_2)$ and available renewable capacity $\mathbf{w} = (w_1, w_2)$, from which hindcast results can be obtained in the obvious way. For compactness of notation, we write $\mathbf{y} = \mathbf{d} - \mathbf{w}$, and again write $\mathbf{z} = \mathbf{x} - \mathbf{y}$.

The Loss of Load Probability (LOLP) and Expected Power unserved (EPU) conditioned to fixed values of demand and wind may be written in general form as:

$$\mathrm{LOLP}_1 = \int f_{X_1}(x_1) f_{X_2}(x_2) \mathbb{I}(d_1 > x_1 + w_1 + g_1(\mathbf{x}, \mathbf{d}, \mathbf{w})) \mathrm{d}\mathbf{x} \tag{3.1}$$

$$\mathrm{EPU}_1 = \int f_{X_1}(x_1) f_{X_2}(x_2) (d_1 - x_1 - w_1 - g_1(\mathbf{x}, \mathbf{d}, \mathbf{w}))_+ \mathrm{d}\mathbf{x} \tag{3.2}$$

where f denotes the corresponding density functions, the indicator function $\mathbb{I}$ is 1 if the condition in brackets is true and 0 otherwise, $(a)_+ = \max(a, 0)$, and g_1 is the capacity available for import to area 1 (which is in general depend on the shortfall-sharing policy and is a function of demand and generation in both systems; see Figure 3.2 for available flows under a veto and share policies). Both (conditional) LOLP and EPU can thus be written as expectations of functions over the range of possible conventional plant availabilities (see Figure 3.3 for the precise integration regions of veto and share policies in available conventional generation space).

3.3.1 Veto policy

The first sharing policy considered is a *veto* power flow policy, in which areas only export whatever available generation surplus they have after satisfying their own demand; this ensures that the contribution of interconnectors is non-negative for

all areas. This calculation has been performed in many papers and textbooks (Billinton and Allan, 1985; Singh et al., 2018), however the way it is stated here is slightly different from most treatments in the adequacy literature.

The LOLP in area 1 can then be calculated by considering the three cases in which the capacity surplus/deficit at area 2 is respectively negative, between 0 and the interconnector capacity c, and greater than c:

$$\begin{aligned}\mathrm{LOLP}_1 = {} & P(X_2 < y_2)P(X_1 < y_1) \\ & + \int_{y_2}^{y_2+c} f_{X_2}(x_2)P(X_1 < y_1 - (x_2 - y_2))\mathrm{d}x_2 \\ & + P(X_2 > y_2 + c)P(X_1 < y_1 - c) \end{aligned} \tag{3.3}$$

in the case of a continuous distribution of available conventional capacity. The three cases above correspond to area 2 having a capacity shortfall (therefore not being able to provide any capacity to area 1), area 2 not having a shortfall but not being able to cover area 1 shortfalls completely, and area 2 being able to saturate the interconnector but area 1 having a shortfall larger than interconnection capacity (see Figure 3.1 for a depiction of the capacity shortfall region in pre-interconnection capacity surplus/deficit space). For calculations presented here, capacities are rounded to the nearest MW and the integral converted to a sum. A similar expression may be derived for EPU in area 2.

3.3.2 Share policy

In this policy, areas share shortfalls in proportion to their demand level, to the extent that the interconnection capacity allows. This could be the result of an agreement between system operators or areas within a system to pool security of supply, and an economic interpretation has also been suggested in (Tindemans et al., 2019) where we believe this policy was first studied.

In a 2-area system with an infinite interconnector capacity, a shortfall occurs when the aggregate capacity surplus/deficit is negative, i.e., $z_1 + z_2 < 0$. Then the surplus/deficit in 1 given transfer over the interconnector would be

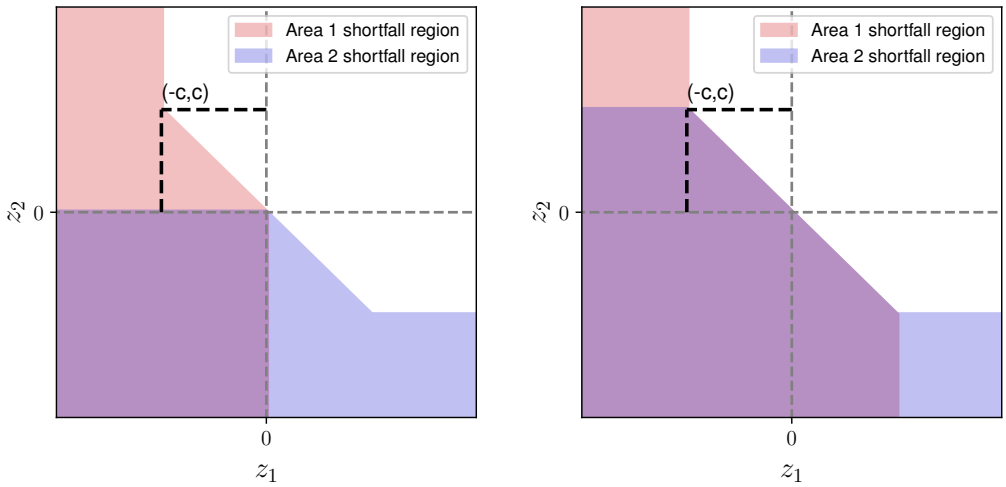
$$z_1^{\infty} = \frac{d_1}{d_1 + d_2}(z_1 + z_2) < 0 \tag{3.4}$$

and the power flow through the interconnector would be the difference between

pre and post interconnection surplus/deficits

$$\delta_1 = z_1^{\infty} - z_1 = \frac{d_1}{d_1 + d_2} z_2 - \frac{d_2}{d_1 + d_2} z_1 \tag{3.5}$$

This transfer depends on both the demands and renewable capacities, and not just on the net demands in each area. It is also notable that under 'share', an area can have a shortfall even if it would not have had one without the interconnector; if both systems have a shortfall, area 1 would still benefit (this is, draw power from area 2) if $z_1/z_2 \leq d_1/d_2$, so that shortfalls are shared in the same proportion as total demand. The finite interconnector capacity is imposed by simply capping the flow at that interconnector capacity.

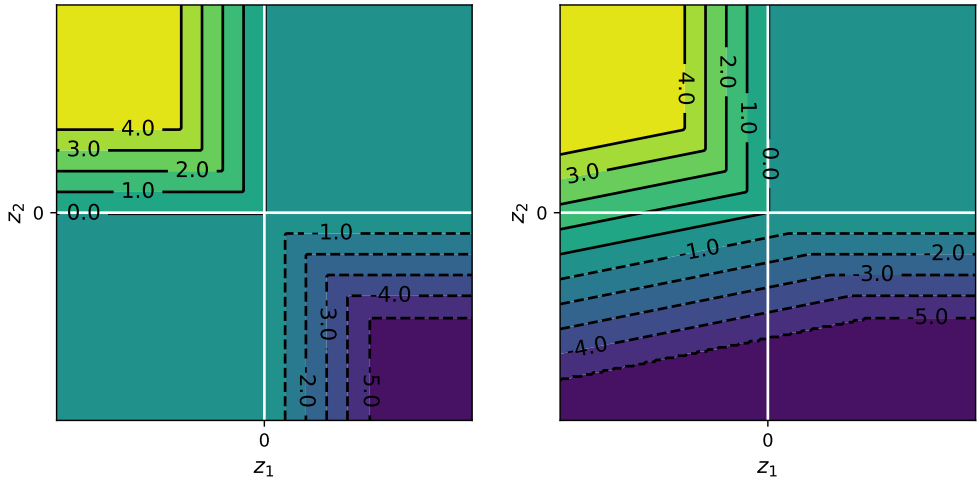


(a) Shortfall regions (veto)

(b) Shortfall regions (share)

Figure 3.1: Shortfall regions under veto and share policies in pre-interconnection capacity surplus/deficit space (z_1, z_2). Overlapping regions indicate capacity shortfall regions for both systems .Note that under a share policy shortfalls can occur even when an area has positive pre-interconnection capacity surplus/deficit depending on capacity shortfalls in area 2.

As for the share policy, there are three cases to be considered in any calculation, namely a flow of size c to 1, a flow of size c from 1, and interconnector flow not on its limit. Closed form expressions may be derived for the boundary in the conventional generation space (x_1, x_2) between these regions as illustrated in Fig. 3.3a, and for each region a formula for the LOLP or EPU in area 1 may be derived in terms of a single integral over x_1 (see Appendix B). While computation has previously been performed for this policy in (Tindemans et al., 2019), we expect the scheme presented here in terms of single integrals to be more efficient than the computational scheme in that previous work, which performed an



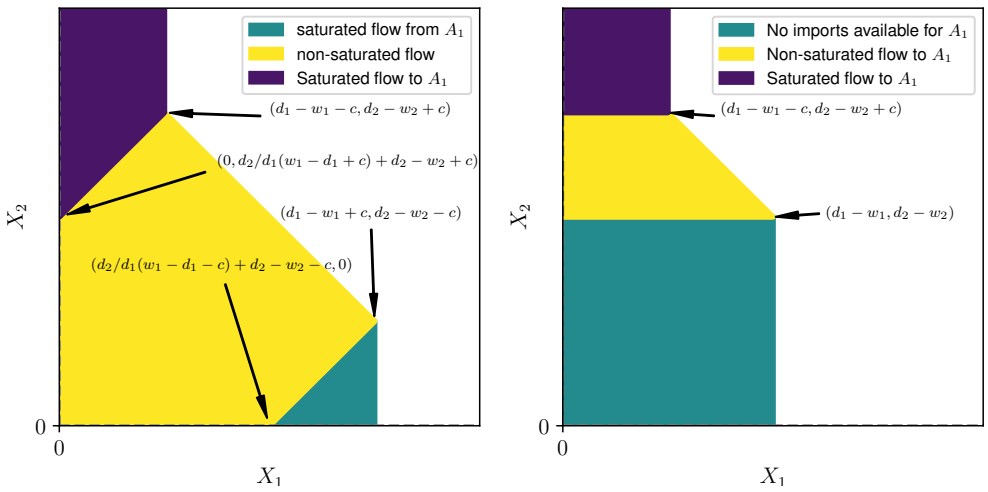
(a) Available imports/exports (veto) (b) Available imports/exports (share)

Figure 3.2: Available imports/exports to area 1 under veto and share policies in pre-interconnection capacity surplus/deficit space (z_1, z_2); dark regions denote exports from area 1, expressed as negative imports. In the case of share, available flows depend on demand proportions at the time of capacity shortfalls too as per Eq. (3.5). For visualisation purposes, demand proportions were fixed to $5/6 \approx 0.83$ for area 1, which is approximately the proportion of GB demand in the GB-IRL system; this ratio determines the angle between the sloped contours and the x-axis in the lower right Figure. Finally, interconnection size was set to 4 GW.

explicit double integral over (x_1, x_2) using fast Fourier transform.

3.3.3 Interconnection EFC

We can value the interconnection's contributions to security of supply in a given 2-area system in terms of its equivalent firm capacity (EFC) for each of the individual systems. Capacity metrics such as EFC allow a direct comparison between system components in terms of the contribution to system adequacy that they make, even when said assets are qualitatively different and are otherwise difficult to compare. In this case for instance, an interconnector makes it possible for imports to flow but do not itself produce any power, which in a sense complicates comparison to traditional generating units. The EFC is defined as the amount of ideal (i.e., fully reliable) capacity to which the system component under consideration is equivalent in terms of system risk. This implicitly assumes a risk indices under which the equivalency is measured. For instance, under LOLE a generating unit's EFC is simply it's average long-term availability times its generating capacity. In the case of interconnection, let $r(\cdot)$ be a risk indices as a function of a system's capacity surplus/deficit distribution; let $Z, \mathcal{Z}$ be the pre



(a) Share policy (b) Veto policy

Figure 3.3: Regions in conventional generation space (x_1, x_2) that cause area 1 to experience a capacity shortfall for fixed values of demand and wind $\boldsymbol{d} = (d_1, d_2), \boldsymbol{w} = (w_1, w_2)$. Different policies are displayed in each figure, and subregions are coloured by interconnection saturation types. For a share policy, the shortfall region is larger due to some shortfalls originating from area 2.

and post-interconnection capacity surplus/deficit of a system respectively, then interconnection's EFC is such that

$$r(Z + \text{EFC}) = r(\mathcal{Z}),\ \text{EFC} \in \mathbb{R} \tag{3.6}$$

The EFC of an interconnection with a given capacity c in general varies depending on the shortfall sharing policy assumed, the chosen risk metric, and the system background to which the interconnector is added. Note Eq. (3.6) valuates interconnection from the point of view of the local system and assumes the neighbouring system does not change (e.g. to take advantage of interconnection's adequacy benefits). Computing simultaneous interconnection valuations for both areas is a more complex problem (See the end of Section 2.6.1 for a brief review of relevant literature in this regard).

3.4 Data

The demand data we use consist of historical hourly measurements from both system operators that have been rescaled using yearly Average Cold Spell estimates

as in reference (National Grid, 2017a). Hourly wind generation measurements are based on preprocessed MERRA data as in (Staffell and Pfenninger, 2016), aggregating estimated wind capacity factors over the locations of wind farms as of January 2015. A total wind generation capacity of 15 GW and 3 GW was used for GB and Ireland, respectively, and different installed wind capacities can be obtained by rescaling these numbers by a constant factor. In the case of availability probabilities and capacities for each generator in each area, we use generation scenarios for 2016-2017 from Baringa Ltd. (Baringa Ltd, 2016) in the case of Ireland, and generation scenarios from National Grid in the case of GB. Finally, demand and wind data were rescaled so as to give an overall LOLE of 3 hours per year when pooling data from 2007 to 2013 inclusive; this was done for both countries without considering interconnection.

In the remainder of this work, we refer to each peak season in the data by the year at which the season started, so for example we refer to the 2007-2008 winter just by 2007.

3.5 Numerical Experiments

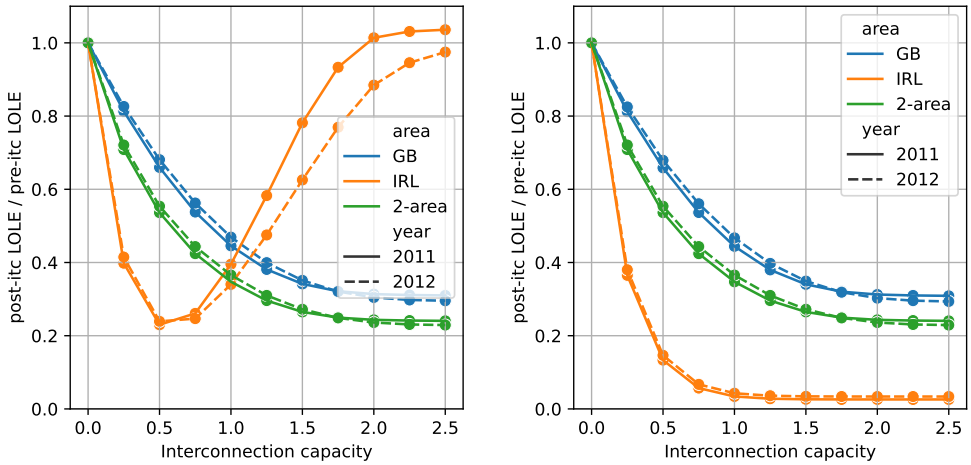
We use the formulation described so far to calculate the value of the interconnection for each area in terms of EFC and risk reduction, taking the isolated version of each system as a baseline. The goal of the experiments described in this section is to assess how interconnection contributions vary under different system configurations, in particular regarding shortfall sharing policies and an increasing proportion of renewable generation. The qualitative behaviour of the EEU under the different system configurations studied was in general similar to LOLE, so for some experiments we will show results for LOLE only.

Year	GB	IRL
2007	0.73	0.79
2008	2.55	3.96
2009	2.79	4.23
2010	8.04	9.38
2011	1.89	0.56
2012	4.90	1.47
2013	0.10	0.62

Table 3.1: Table of baseline LOLE values expressed in hours per years, to which we compare risk indices for different system configurations. The values shown here correspond to a null interconnection capacity for data from the corresponding year.

3.5.1 Interconnection under different policies

The qualitative behaviour of the LOLE and EEU as we varied interconnection capacity was fairly consistent across years for both areas. In Fig. 3.4 we show a representative comparison of 2011 and 2012 for a range of interconnections capacities (X-axis). In this plot, we show how the LOLE evolves as we add more interconnection capacity, and we do this for each combination of year, area and policy. The LOLE for each year is shown as a proportion of the baseline value for that year – these baseline values can be found in Table 3.1.



(a) Share policy

(b) Veto policy

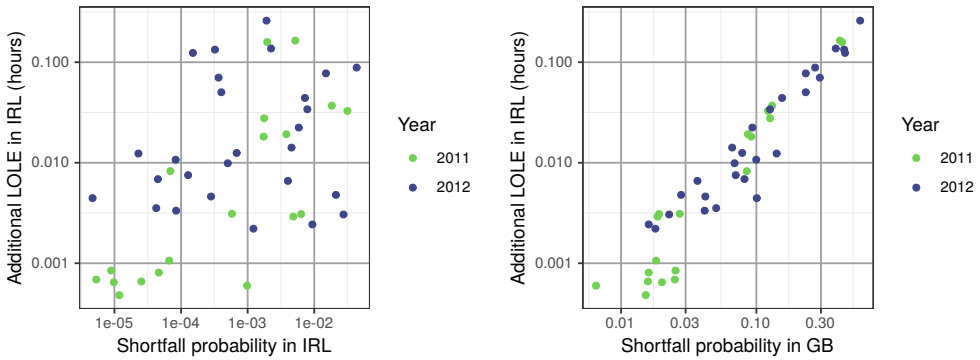
Figure 3.4: Traces of LOLE as proportions of their baseline values in table 3.1. For each combination of year, area and policy, we show how LOLE evolves as we add interconnection capacity. Interconnection capacity is on the X axis, while the Y axis shows post-interconnection LOLE as a percentage of pre-interconnection LOLE.

Much of the sensitivity of risk to different parameters was consistent across all years; in Fig. 3.4 we show 2011 and 2012, years of low and above-average risk for GB respectively, and low risk for IRL. In the case of GB, both policies entail practically the same decrease in risk for all interconnection capacities and years; interconnection decreases risk to around 30% of that of an isolated GB system, and its contributions plateau between 1 and 2GW capacity; all of these behaviours were consistent across system configurations.

Risk in Ireland consistently becomes negligible under a veto policy above 1 GW capacity. However, it was found that under a share policy, additional interconnection capacity starts increasing the shortfall risk after 0.6GW; this effect was also consistent across historic years but particularly strong in 2011 and 2012; in the latter, the interconnection contribution becomes slightly negative

at around 2.5GW. The same effect is observed for EEU, although the increase in risk stabilises at around 50% of the baseline risk value. This implies that a share policy could make relatively small scale shortfalls more likely in Ireland. The underlying reason for this behaviour is the asymmetry in the ability of the system to provide sufficient surplus to cover a shortfall in the other area, due to the smaller size of the Irish system.

To show this, we use data from 2011 and 2012 to find times in the historic record that are driving most of the increase in Irish LOLE when going from 0.6 to 2.5GW interconnection under a share policy. In Figures 3.5a and 3.5b we show these times and the corresponding increase in LOLE they induce in Ireland (Y-axis). In these figures, the scale we use for the x-axis is the shortfall probability that each of these net demand observations induce in the corresponding area, which makes it straightforward to compare how shortfall risks in each of the two areas drive the Irish increase in LOLE. It is apparent from looking at both plots that what drives almost all of the increase in Irish LOLE when adding interconnection capacity is the shortfall probability in the GB system and not so much the Irish system itself; there seems to be a strong linear relationship between the logarithms of the Irish LOLE increase and shortfall probabilities in GB. The correlation coefficient between the GB shortfall probability and the Irish LOLE increase is 0.95 – this reduces to 0.56 when comparing the Irish shortfall probabilities and LOLE increases.



(a) Highest risk-inducing observations from the IRL system

(b) Highest risk-inducing observations from the GB system

Figure 3.5: Observations from 2011 and 2012 in each one of the systems that cause the largest increase in risk for Ireland when going from 0.6GW to 2.5GW interconnection capacity under a share policy. The Y axis is the net increase in LOLE (in log scale), while the X axis is the shortfall probability that each level of net demand induces in the corresponding system.

3.5.2 Interconnection and renewables integration

Another important factor to consider in this system is the wind penetration factors. Under some of the scenarios considered by National Grid, renewable generation could represent up to 80% of total electricity output (NGESO, 2017) in 2030. We thus analyse how the EFC of interconnection changes as the proportion of wind generation increases in each system's generation mix. To this end, we rescale wind generation data using progressively larger rescaling factors (from the baseline levels of 3GW and 15GW for Ireland and GB respectively), at the same time shifting the distribution of available conventional capacity so as to keep the baseline risk levels fixed throughout the experiments. This ensures we analyse credible system configurations in terms of the current adequacy standards in GB. While any such experiment is, while instructive, somewhat artificial, we judge that shifting the conventional generation distribution is preferable to rescaling demand, as the latter could distort outcomes of the share policy.

We display results for an interconnection capacity of 1GW, i.e., the present interconnector capacity. The value of interconnection is expressed as an EFC, based on pooling historic data from all years. Calculations are performed for wind capacities from 15 to 30 GW and from 3 to 6 GW in GB and Ireland respectively. The results for GB are shown in Figure 3.6 and for IRL in Figure 3.7, for all four possible combinations of risk index (LOLE and EEU) and interconnector policy (veto and share).

In all combinations of policy-metric-area, larger amounts of installed wind capacity in the two systems increase interconnection's EFC. The change in EFC as a function of wind capacities for GB seem to follow a similar pattern in all cases except for a share-EEU combination; in the latter case, the EFC seem to depend almost solely on the Irish installed wind capacity.

Conversely, in the case of the Irish system, EFC seem to depend almost solely on the installed wind capacity in GB when using a share policy regardless of the risk metric used, while under a veto policy sensitivity to installed wind capacity at both areas is more similar. However, for both areas sensitivity of EFC to installed wind capacities is overall small compared to sensitivity to the choice of policy and risk metric.

While in GB, the EFC does not depend too strongly on any of the factors varied and stays in a range of 600-690 MW for all policy-metric combinations, in Ireland the difference can be as large as 250 MW, namely between the veto-LOLE and share-LOLE combinations. Moreover, the choice of risk metric within a given policy for IRL can change interconnection's EFC by roughly 100 MW.

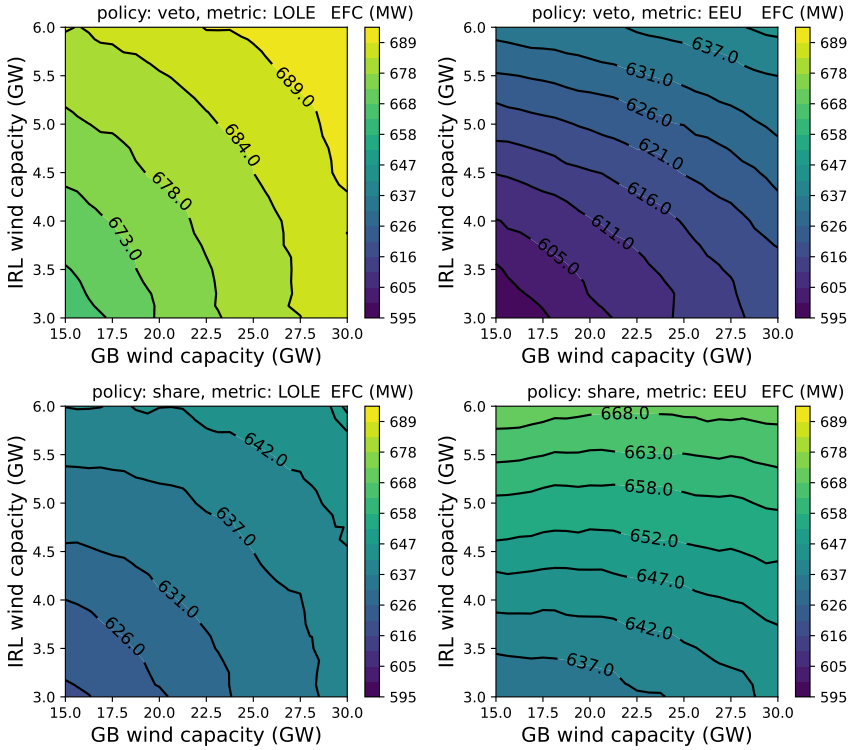


Figure 3.6: Interconnection EFC's (expressed in MW) contour line plots for GB, under different combinations of policy and underlying risk metric.

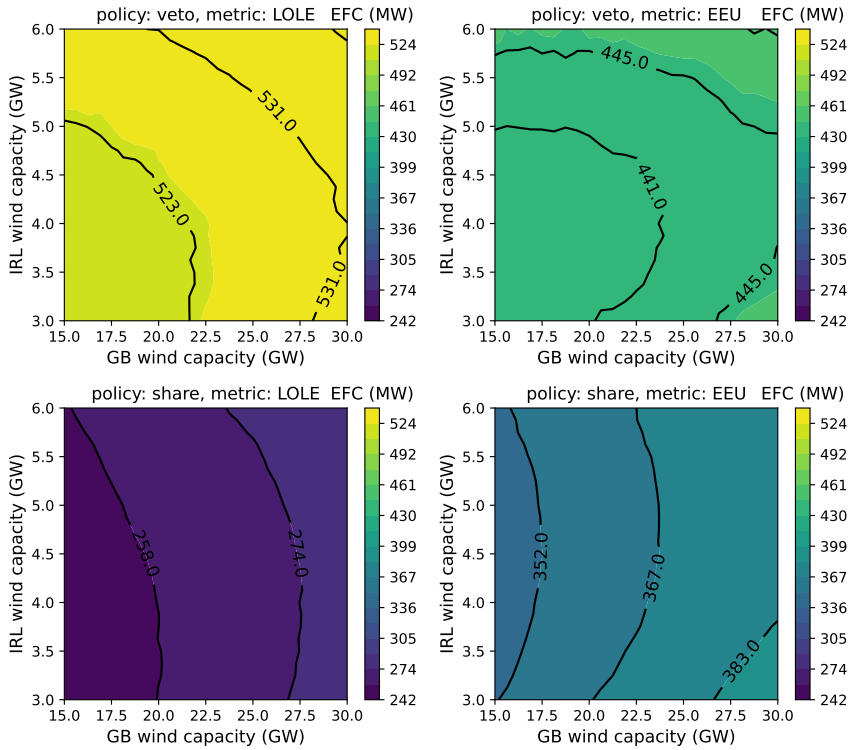


Figure 3.7: Interconnection EFC's (expressed in MW) contour line plots for IRL, under different combinations of policy and underlying risk metric.

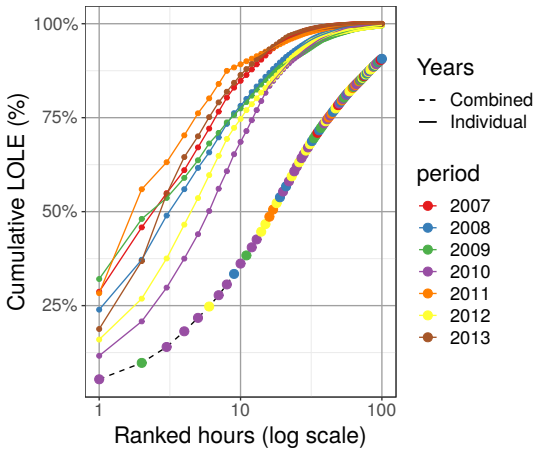


Figure 3.8: LOLE arising from the n historical records with highest (demand-wind) for GB, against n. Only the top 100 observations are shown.

This illustrates the importance of choosing a risk metric which is relevant to the question at hand, and of considering governance arrangements when planning interconnection between systems. Overall, the observation that the range of values of interconnection's EFC remains small given a metric-policy combination might be partly due to dependence between the outcomes of (demand minus renewable generation) in the two areas.

3.6 Future extensions

The risk calculations in this Chapter have used the 'hindcast' approach, i.e., using the empirical historical distribution of (demand - wind) observations as the estimated distribution required for predictive risk calculations. This provides a simple way of accounting for statistical association between wind availability and demand as seen in the historic data, but hindcast results can be driven directly by a very limited number of historic records. Fig. 3.8 illustrates this for GB.

For calculations conditional on a specific historic year of data, around 75% to 90% of the calculated LOLE arises from just 10 historic records, and pooling all the data the equivalent figure is 40%. It is notable however that 9 of the 12 highest-risk records come from a limited period in winter 2010-11, and that more generally given a different prevalence of historic winters in a similar size dataset quite a different LOLE result might be found (arising from sampling uncertainty) even given the same underlying power system and climate. It would thus be valuable to develop the necessary multivariate extreme value approaches to smooth the tail of the distribution of net demand, in analogy to (Wilson and Zachary, 2019) for a single area system.

3.7 Conclusion

This chapter has presented an analysis of how interconnection contributes to system reliability in a 2-area model of the combined IRL-GB system. We demonstrate how the benefit of interconnection to a system can depend on the policy agreed for sharing resource between systems. If a system covers its own demand before making capacity available across the interconnector, then the presence of interconnection can only reduce the level of risk in the model in comparison to the counterfactual of no interconnection. However, for a policy where shortfalls are shared between the two systems in proportion to demand, the presence of interconnection can actually increase the level of risk in the smaller Irish system.

3.8 Discussion

Even though a share policy would, in principle, be compatible with objectives outlined by the European Union regarding the integration of energy networks (Rumpf and Bjørnebye, 2019), results we observed may make it a difficult choice for some (possibly risk-averse) system operators, at least in the form we have explored here, as small systems could potentially experience an increase in overall risk in certain circumstances. As mentioned in (Tindemans et al., 2019), a share policy can be used instead as a conservative estimate on the benefits of interconnection, although the results obtained here suggest this bound might be conservative in some cases.

Experiments regarding valuation of interconnection suggest the choice of resource sharing policies between systems and the choice of risk indices under which interconnection contributions are measured can have a large impact on its valuation. This is an important consideration, for instance, when deciding to what extent interconnectors can participate in capacity markets, as pointed out in (Tindemans et al., 2019). Installed wind capacity seem to have only a small impact on this, and in the GB-IRL system this might be because of strong statistical dependence in wind generation across both areas. Verifying whether this holds true for other systems could be an interesting exercise.

We have also derived efficient, exact calculations for EEU and LOLE risk indices at full resolution in a two-area power systems under the versions of a veto and share policies we use in this work (see Appendix B). Moreover, the implementations are available in the `riskmodels` Python package (Sanchez, 2022) (see Appendix A). These calculations rely on a hindcast model of net demand under a time-collapsed framework (i.e, using the empirical historical distribution

of demand minus wind generation as the estimated distribution for predictive risk calculations), which is similar to the methodology currently used for single-area adequacy calculations in GB. However, it is unlikely this can be extended beyond two areas, as formulas for exact calculations quickly become unwieldy and the involved computation time grows exponentially with the number of systems. Thus for more than two areas, Monte Carlo estimation may be the only viable alternative to compute analogous results when working at full resolution (i.e., without binning probability distributions to a coarser resolution).

Figure 3.8 suggests estimates from hindcast models are dominated by a very small number of points. Although the provided example illustrates this for a single-area system, this issue is likely more serious for two-area or multi-area systems, as there will be an even lower number of observations where net demand is simultaneously high at multiple systems. Moreover, hindcast models cannot account for the possibility of record-breaking net demand values beyond what has been observed in the past, which can certainly occur in a future year. Using results from multivariate extreme value theory to develop a principled alternative to the bivariate hindcast model we have used here is a natural extension to this work, along the direction set by (Wilson and Zachary, 2019). We explore this alternative in the next Chapter.

Chapter 4

Analysing the effects of statistical dependence on the security of supply of the Ireland - Great Britain power system

Adequacy risks originate at the tails of the involved distribution: a combination of unusually high net demand (defining it as demand minus available renewable generation) and unusually low available conventional capacity. Relevant data in the historic record is thus sparse. As a consequence of this sparsity, *hindcast* models of net demand (which use the empirical distribution of forward-mapped historic net demand data as the predictive net demand distribution for risk calculations) like the ones we used in the previous Chapter tend to produce risk estimates which are driven by a very small number of historic observations (see Figure 3.8); this issue is likely more serious in multi-area systems, as the number of points in the joint tails may be even smaller.

In this Chapter, we use results from statistical extreme value theory (EVT) to find smooth, bivariate models for the tails of net demand in a two-area system and to model dependence in capacity surpluses/deficits across areas, using data from the Ireland - Great Britain (IRL-GB) system. In addition to providing parametric models to smooth out the sparse data regime at the tails, the concept of asymptotic dependence from EVT provides a useful theoretical tool for characterising dependence between extreme occurrences. We find strong evidence that the capacity deficits in the GB and Ireland systems are asymptotically independent, despite finding some evidence of asymptotic dependence in net demands across the two areas. This is consistent with the intuition that, when indepen-

dent distributions of conventional capacity are convolved with those of demand and wind, the dependence is considerably weakened. The consequent use of a Gaussian copula to describe the dependence of capacity deficit occurrences provides a convenient means of carrying out sensitivity analysis to the strength of relationship.

The contributions of this chapter are methodological and of a practical nature: we use existing results from the theory of extreme values as well as models from the power system adequacy literature to propose a novel methodology for computing risk estimates in a two-area power system. We use data from the GB-IRL system to demonstrate this methodology and its advantages and compare the results to existing risk quantification models.

4.1 Introduction

Security of supply is one of the three pillars of energy policy, along with affordability and sustainability. In Great Britain (GB), annual capacity auctions take place to ensure that capacity meets reliability standards based on risk estimates for future years (NGESO, 2019). For this reason, risk calculations and their statistical methodology are of considerable interest for policy makers. Such estimates need to consider the existence of interconnectors to other power systems, which potentially represent an additional source of available capacity. This is particularly relevant in the context of decarbonisation, as the aggregation of renewable generation over wider geographical regions is expected to result in less variable power output. On top of this, there are other benefits regarding market integration (European Comission, 2016) like economies of scale from shared planning over wide regions.

Multi-area reliability has been studied since at least the 1960s, and models proposed in the literature for adequacy calculations in multi-area systems are numerous (see Sections 2.6 for a brief overview). This variety is reflected in the range of practices found across different system operators. National Grid conducted an international survey in 2017 (National Grid, 2017b) and it was found, for instance, that some system operators such as those in France and Belgium as well as some parts of the US, have relatively fine-grained market simulation tools for neighbouring systems and use them in their reliability studies, while others take available imports as fixed. In the case of Great Britain, an offline market model is also used to specify a distribution of available flows through interconnectors conditioned on different demand percentiles, and this distribution

is convolved with the overall supply balance thereafter. We note that the use of market modelling tools in adequacy calculations could be problematic, as market simulation tools assume (implicitly or explicitly) regular operating conditions in their models, whereas adequacy risks are almost entirely determined by the occurrence of extreme events for which assumed conditions might simply not hold.

Most of the risk regarding capacity shortfalls come from rare events either at the demand side, which can be unusually high perhaps due to cold weather in the case of Great Britain (GB), or at the generation side, where conventional generators can fail unexpectedly and renewable generation can experience sustained drops in generation output due to certain weather patterns. Interconnected systems are usually geographically close, and so their demand and renewable generation output is likely to be statistically dependent through the action of weather systems and similar energy usage patterns. Simultaneous occurrence of extreme demand and renewable generation in both areas can have a large effect on the value of interconnection in terms of security of supply. Under strong statistical dependence, it is more likely that both systems experience stress at the same time, with low availability of imports or exports as a consequence.

It is a common practice in reliability studies both in industry and in the literature to use the empirical distribution of historic load as input, either suitably rescaled or as is (National Grid, 2017b; Sheehy et al., 2016; Billinton and Allan, 1985), as the predictive distribution for risk calculations. Chapter 3 uses a model like this to value interconnection in the IRL-GB system. The tails of empirical distributions are by definition sparse, with all of the probability mass in these regions concentrated in only a handful of points representing previously observed extreme levels. As risk comes almost entirely from the tails, a consequence of this is that risk indices can be essentially determined by a very small set of past observations (Sanchez et al., 2020). Moreover these models implicitly assume that no record-breaking events can occur in the future. Results from EVT provide a sound methodological basis for finding smooth parametric alternatives to the sparse tail regimes of hindcast models, and to characterise the tails of the involved distribution and the dependence between them (i.e., extremal dependence) with a small number of interpretable parameters. This in turn makes it straightforward to perform a sensitivity analysis of interconnection value to the type and strength of statistical dependence. Models from EVT have been used before in a single-area system model using data for GB (Wilson and Zachary, 2019); we build on this work by developing appropriate statistical dependence models between the two areas of the IRL-GB system.

Models of statistical dependence (e.g. copula models) have been used before

in the adequacy literature, particularly for modelling wind generation. For example, in (Hagspiel et al., 2012), copulas are used to model dependence in wind generation across the European network, using the resulting model to examine the effects of increasing wind penetration on the Swiss power grid. The use of copulas to model wind generation at smaller geographical scales, e.g. across wind farm sites or even across individual wind turbines, has also been studied (Hu and Bo, 2013; D'Amico et al., 2015). However, to our knowledge, extreme value copulas and other results from EVT regarding statistical dependence have not been explored in multi-area adequacy calculations.

The outline of the chapter is as follows: in Section 4.2, we give a brief review of standard power system risk models and indices from the literature that we use in this work. In Section 4.3, motivation for the use of extreme value theory is given, and the most relevant results for this work are briefly discussed, referencing more comprehensive material on the subject; Section 4.4 describes the data we used. In Section 4.5 we describe the methodology for our two-area net demand model, which is defined as demand minus renewables, and its comparison to a hindcast net-demand model; in Section 4.6 we switch focus to analysing statistical dependence in capacity surpluses, characterising its dependence and performing sensitivity analysis of different risk indices to statistical dependence. Finally, Section 4.7 outlines the Chapter's conclusions.

4.2 System model

In this section, we give a brief overview of some of the existing models from the literature of security of supply that will be using in this work. We denote random variables with uppercase letters and constants with lowercase letters. Vectors will use bold letters. We use X, W and D to denote random variables corresponding to available conventional generation, renewable generation and demand respectively. Furthermore, we define net demand, or demand net of renewables, as $Y = D - W$, and capacity surplus as $Z = X - Y = X - (D - W)$.

4.2.1 Single area system

Let the period of interest (say, a peak season) be divided into n segments Δ_t, usually expressed in hours. We will consider two risk indices, the loss of load expectation (LOLE), which is the expected number of hourly[1] shortfalls in the

[1]For simplicity, all results presented here are for an hourly time step, as per the available data from GB – in N American terminology this corresponds to LOLH. The results generalise

period, defined as

$$\text{LOLE} = \mathbb{E}\left[\sum_{t=1}^{n} \mathbb{I}(Z_t < 0) \cdot \Delta_t\right] = \sum_{t=1}^{n} \mathbb{P}(X_t < Y_t) \cdot \Delta_t, \tag{4.1}$$

and the expected energy unserved (EEU), which is the expected amount of energy not supplied, defined as

$$\text{EEU} = \mathbb{E}\left[\sum_{t=1}^{n} \max\{0, -Z_t\} \cdot \Delta_t\right]. \tag{4.2}$$

Here t indexes times in the future season or year under study. For the purposes of statistical modelling, as in this Chapter, it is often more convenient to work in a *time-collapsed* picture with the time-collapsed variable Z representing surplus at a randomly chosen point in time in the peak season under study. For a system such as that considered here, which does not have storage or other technologies which link time periods in a similar way, the LOLE is then calculated as

$$\text{LOLE} = n \cdot \mathbb{P}(Z < 0) = n \cdot \mathbb{P}(X < Y), \tag{4.3}$$

and an analogous formula applies for EEU.

The most common means of estimating the distribution of net demand is to use the forward-mapped empirical historic distribution as the distribution for predictive risk calculations, sometimes referred to as *hindcast* (Zachary and Dent, 2014; Keane et al., 2011). The distribution of net demand is then given by

$$\mathbb{P}(D - W \leq y) = \frac{1}{T} \sum_{\tau=1}^{T} \mathbb{I}(d_\tau - w_\tau \leq y). \tag{4.4}$$

where T is the number of observations in the historic record, τ indexes the historic records, and historic demand and wind resource have been appropriately rescaled to the future system scenario under study. This approach may also be interpreted as estimating the risk conditional on a repeat of historic conditions in one or more years.

Available conventional generation X is modelled in a time-collapsed fashion as well, as outlined in Section 2.3.1, using historic data on outage rates. Within the examples presented, risk indices that depend on the distribution of Z may then be calculated by convolving the distributions of X and Y. However, a time-sequential model would in general be needed to consider technologies such as

in a straightforward way to calculations for higher time resolutions.

storage that link time periods in the calculations, even when only expected value indices such as EEU and LOLE are evaluated.

4.2.2 2-area system

In a 2-area system, interconnection can make imports available to a system under stress, up to the interconnection capacity, thereby reducing risk for both systems. The impact of interconnection on the risk indices depends not only on interconnection capacity and statistical dependence across areas, but also on the shortfall-sharing policy, understood here as an agreement of how the interconnector is used when one or both systems are under stress. Various policies have been studied in the literature. For instance, (Tindemans et al., 2019) discusses so-called *veto* and *share* policies, the former being that in which a power system exports only spare available capacity, if any, while in the latter case shortfalls are shared across areas according to an agreed rule. In the rest of this Chapter, we assume a *veto* policy between both areas, as this will suffice to illustrate the statistical approaches developed. If $\boldsymbol{Z} \in \mathbb{R}^2$ is the bivariate pre-interconnection vector of capacity surpluses/deficits across both systems, and there is an interconnector with capacity $c \geq 0$ between them, the post-interconnection shortfall region for area 1 under a veto policy can be divided in two subregions:

$$\mathcal{R}_1 = \{\boldsymbol{Z} \in \mathbb{R}^2 \,|\, Z_1 < -c\} \tag{4.5}$$

$$\mathcal{R}_2 = \{\boldsymbol{Z} \in \mathbb{R}^2 \,|\, -c \leq Z_1 \leq 0, Z_2 < -Z_1\} \tag{4.6}$$

which represent the cases where area 1 has a shortfall larger than interconnection capacity, and the case of area 2 not being able to cover a shortfall in area 1, despite the shortfall being smaller than interconnection capacity (see Figure 3.1a for a depiction of these regions).

Under a snapshot model the two risk indices may be calculated as follows for area 1:

$$\text{LOLE} = n \cdot \mathbb{P}(\boldsymbol{Z} \in \mathcal{R}_1 \cup \mathcal{R}_2) \tag{4.7}$$

$$\text{EEU} = n \cdot \mathbb{E}[-Z_1 \cdot \mathbb{I}(\boldsymbol{Z} \in \mathcal{R}_1 \cup \mathcal{R}_2)] \tag{4.8}$$

where $\mathbb{I}$ is an indicator function. This is, there is only a shortfall in area 1 when the pre-interconnection surplus vector $\boldsymbol{Z}$ is in $\mathcal{R}_1 \cup \mathcal{R}_2$. For area 2, the reasoning is analogous with reversed indices.

4.3 Extreme value theory

In terms of LOLE and EEU, virtually all of the risk is concentrated in the tails of net demand, that is, it comes from the highest net demand values. As mentioned before, this concentration is especially severe in empirical hindcast models of net demand, where risk indices can be determined by a very small number of historic records with the highest net demand observations. In order to alleviate this problem, we turn to EVT-based models which provide a smooth parametric alternative to hindcast models.

EVT is a branch of statistics that provides mathematically principled methods for making inferences about statistical properties of extreme events, including those rare enough to be outside the historic record's range. These models arise as limiting behaviour of sample maxima or exceedances above progressively large levels. EVT makes only mild regularity assumptions on the data while providing general results on extreme occurrences of random variables. It is widely applicable and routinely used in fields from insurance to environmental sciences (Beirlant et al., 2006; Coles, 2013). In this work we are interested in the threshold-exceedance framework which we outline below. Then, we briefly discuss relevant results in multivariate EVT.

The following section provides a brief overview of the concepts from EVT that are relevant to this Chapter. We emphasise the reviewed content is not new, and is in fact well known in the EVT community. For a more detailed (but still concise) overview of these methods and concepts, see Appendix C.2; a full treatment of these topics can be found at (Beirlant et al., 2006; Coles, 2013).

4.3.1 Univariate exceedances

A key result from EVT states that under mild assumptions on the distribution of a random variable $X \sim F(x)$, exceedances over a threshold μ, given by the conditional distribution $X\,|X > \mu$, follow a Generalised Pareto distribution (GPD) in the limit, as $\mu \to \infty$ (Coles, 2013). For the purpose of this work, this result means that for an appropriately large threshold μ, net demand exceedances over said threshold, conditioned on $D - Y > \mu$, are well approximated by a GPD, whose cumulative distribution function is given by

$$\mathbb{P}(W \leq w) = 1 - \left(1 + \xi \left(\frac{w-\mu}{\sigma}\right)\right)_{+}^{-1/\xi}, w > \mu \tag{4.9}$$

where $\xi \in \mathbb{R}$ and $\sigma > 0$, and reduces to an exponential distribution if $\xi = 0$.

A semiparametric model for the full data range can then be constructed by using the fitted GPD model for tail exceedances above the chosen threshold μ, and the empirical data distribution $\hat{F}$ below it. The full model can thus be written as

$$\bar{F}(Z \leq z) = \begin{cases} \hat{F}(z) & z \leq \mu \\ \hat{F}(\mu) + (1 - \hat{F}(\mu)) * F_{GP}(z) & z > \mu \end{cases} \tag{4.10}$$

where F_{GP} is the fitted GPD. A model like this has already been applied to a single-area power system using data from GB (Wilson and Zachary, 2019).

4.3.2 Asymptotic dependence and multivariate EVT

When dealing with extremes of multiple random variables, as is the case in a two-area system, EVT also provides a framework to measure the degree of association between different components at extreme levels, which in our problem directly influences the utility of interconnection in terms of security of supply.

A central concept in this context is asymptotic dependence, which quantifies the degree to which extreme values of different components occur together. Considering that the choice of marginal distributions does not affect the dependence structure between random variables, and that the marginals can be chosen by applying a particular transformation to each random variable, we assume without loss of generality that Y_1, Y_2 are standard Frechet random variables with marginal CDF $F(y) = \exp(-1/y), y > 0$. Then, we define the χ statistic as

$$\chi = \lim_{y \to \infty} \mathbb{P}\left(Y_2 > t \,|\, Y_1 > y\right) \tag{4.11}$$

The variables are asymptotically dependent if $\chi > 0$, and asymptotically independent otherwise. Asymptotically independent variables can still exhibit strong dependence at non-extreme levels, with dependence vanishing only in the tails. An example of this is the bivariate normal distribution with correlation $-1 < \rho < 1$ whose components can be proven to be asymptotically independent, regardless of ρ (Beirlant et al., 2006, p. 285). Determining the presence of asymptotic dependence is important in devising an appropriate model for the data. There are many well studied parametric models for asymptotically dependent data, while asymptotically independent data may require semiparametric or non-parametric modelling approaches instead, as done in (Heffernan and Tawn, 2004).

Although a useful theoretical concept, precise estimation of χ is not straightforward, and typically visual inspection of empirical approximations are used to

assess asymptotic dependence. A related statistic that is more amenable to numerical estimation is the coefficient of tail dependence η (Beirlant et al., 2006, p. 345); this is bounded by $0 \leq \eta \leq 1$, and is defined by making an additional assumption in the context of (4.11), namely

$$\mathbb{P}(Y_1 > y, Y_2 > y) = \mathcal{L}(y) \cdot \mathbb{P}(Y_1 > y)^{1/\eta}, y > 0 \tag{4.12}$$

where $\mathcal{L}(x)$ is a slowly varying function, that is, $\mathcal{L}(yz)/\mathcal{L}(y) \to 1$ as $y \to \infty$ for all $z > 0$. Intuitively, slowly varying functions become flat rapidly (in a precise sense) as $y \to \infty$ (note that $\mathcal{L}$ does not necessarily converge, e.g. $\mathcal{L}(z) = \ln(z)$). This class of functions arise in many important results from the theory of extreme values, and (4.12) has been shown to be valid for a broad range of conditions and models (Ledford and Tawn, 1997; Heffernan, 2000). Here, η describes the type of dependence and $\mathcal{L}(y)$ its strength within the dependence type given by η. We have $\eta = 1$ whenever $\chi > 0$, and $0 \leq \eta < 1$ otherwise, thus characterising the presence of asymptotic dependence; furthermore, we can estimate η by maximum likelihood as the shape parameter ξ in (4.9) from the transformation $Z = \min\{Y_1, Y_2\}$. In practical terms, this allows us to estimate η for our net demand data $y_1, ..., y_t$ by transforming it to approximate standard Frechet margins using the transformation $v_i = 1/(1 - \bar{F}(y_i)), i = 1, ..., n$ for the net demand samples of each individual system, with $\bar{F}$ as in (4.10), and using the component-wise minima of the transformed data $z_i = \min\{v_i^{(1)}, v_i^{(2)}\}, i = 1, ..., n$, where superscripts correspond to each of the systems. Moreover, likelihood estimation of η allows for measuring estimation uncertainty, which will be useful later in this work.

4.3.3 Models of extremal dependence

Models of statistical dependence can be described using copula functions. A copula is a multivariate distribution with uniform marginals in $[\mathbf{0}, \mathbf{1}]$, and can be constructed for any given distribution simply by transforming its marginals to standard uniform. Sklar's theorem states that any multivariate distribution can be written in terms of its univariate marginal distribution functions and a copula function (Sklar, 1959).

Copulas arising from dependence between extremes are called extreme value copulas. More specifically, $C^*(\boldsymbol{u})$ is an extreme value copula if there is a copula $C(\boldsymbol{u})$ such that

$$C^*(\boldsymbol{u}) = \lim_{n \to \infty} \left(C(\boldsymbol{u}^{1/n})\right)^n \tag{4.13}$$

This means the copula between component-wise sample maxima from $C(\cdot)$ con-

verges to $C^*(\cdot)$ as the sample size goes to infinity, and we say that $C(\cdot)$ is in the *domain of attraction* of $C^*(\cdot)$. Sometimes it is more convenient to characterise extreme value copulas through their *Pickands dependence function* $A(t), 0 \leq t \leq 1$ (Beirlant et al., 2006, p. 285). This is a convex function bounded by $\max\{t, 1-t\} \leq A(t) \leq 1$, and has a one to one relationship with $C^*(\boldsymbol{u})$. For its precise definition and relationship to extreme value distributions, see Section C.2 and Definition 6 in the Appendix. In the two-dimensional case any function with these characteristics induces an extreme value copula, and we can write

$$C(u, v) = (uv)^{A(\log(v)/\log(uv))} \tag{4.14}$$

From the above we can see that $A(t) = 1$ produces independent components. Conversely, $A(t) = \max\{t, 1-t\}$ produces perfectly dependent components.

One of the simplest parametric model of extremal dependence is the so called *logistic* model, defined by

$$A(t) = \left(t^{1/\alpha} + (1-t)^{1/\alpha}\right)^{\alpha}, \ \alpha \in (0, 1] \tag{4.15}$$

This Pickands function induces a Gumbel-Hougaard copula, given by

$$C(u, v) = \exp\left(-\left((-\log u)^{1/\alpha} + (-\log v)^{1/\alpha}\right)^{\alpha}\right), \ \alpha \in (0, 1] \tag{4.16}$$

which has been shown to arise as an extremal dependence structure from a wide family of copulas in the underlying distributions (Gudendorf and Segers, 2010). However, from (4.15) we see that this model entails symmetry in $A(t)$ in the sense that $A(1/2-t) = A(1/2+t)$ for all $0 \leq t \leq 1/2$, which is not always appropriate. A more flexible generalisation of this model is the *asymmetric logistic model* whose Pickands function in the bivariate case is given by

$$A(t) = (\psi_2 - \psi_1)t - \psi_2 + 1 + \left((\psi_1 t)^{1/\alpha} + (\psi_2(1-t))^{1/\alpha}\right)^{\alpha} \tag{4.17}$$

with $0 \leq \psi_1, \psi_2, \alpha \leq 1$. This reduces to the logistic model if $\psi_1 = \psi_2 = 1$, and to independence if $\alpha = 1$ or $\psi_1 = 0$ or $\psi_2 = 0$.

Another advantage of working with Pickands functions is that they provide a way to visually inspect the goodness of fit of a model for extremal dependence by comparing the fitted model's Pickands function to the empirical approximation induced by the data, as in Fig. 4.4.

Note that most if not all parametric models of extremal dependence make the assumption of asymptotic dependence. This is because the only possible

limiting behaviour from asymptotically independent copulas is full independence. This means that for, say, normally distributed data with correlation $-1 < \rho < 1$, the dependence between components weakens at progressively more extreme levels and disappears completely in the limit, which severely reduces the utility of parametric, asymptotic approximations in the case of asymptotically independent data. However, many useful semiparametric approaches have been explored; see (Heffernan and Tawn, 2004).

4.4 Data

Wind data were obtained from (Staffell and Pfenninger, 2016). They consist of hourly wind capacity factors based on atmospheric reanalysis data and the locations of installed generation on January 2015, and for this work it has been rescaled to a total wind capacity of 3 GW in IRL and 15 GW in GB. Different installed wind capacities can be obtained for purpose of numerical experimentation by rescaling by a constant factor.

Demand data consist of hourly measurements for the peak seasons of 2007 to 2013 for both systems; GB demand data was obtained from (Staffell and Pfenninger, 2018) while data for IRL was provided by Baringa Ltd. The data has been standardised by rescaling each peak season by their corresponding Average Cold Spell estimates as in reference (National Grid, 2017a) to correct for external factors such as economic growth but preserving variation due to weather patterns (Wilson and Zachary, 2019). Subsequently, all normalised peak seasons were again rescaled to fix the average seven-year period LOLE to three hours per year.

Capacity and failure probability data for conventional generating units in Ireland were developed by Baringa Ltd for the Irish Single Electricity Market Committee in 2016, and are publicly available at (Baringa Ltd, 2016). In the case of Great Britain, data were provided by National Grid, and were anonymised to protect its sensitivity; we take the anonymised data as representative of the real system.

Finally, in this work, we refer to each historic peak season in the data by the year at which the season started, so for example we refer to the 2007-2008 winter just by 2007. We illustrate results on a subset of years in the body of the Chapter; the full series of plots for all years is shown in Appendix C.1.

4.5 Modelling net demand

As mentioned above, we assume that available conventional generation is independent between the systems, and so statistical association in capacity surpluses or deficits comes entirely from net demand across areas. The main motivation in looking for a smooth alternative to a hindcast net demand model is its tendency to produce risk estimates that are almost entirely determined by a very small number of points. Figure 4.1 shows the concentration of LOLE in the highest net demand observations for each season under a hindcast model. For instance, for 2011 in GB roughly 80% of the estimated pre-interconnection LOLE comes from just eight observations. Moreover, this concentration is exacerbated in post-interconnector calculations, particularly for IRL which is smaller relative to interconnection size. The use of smooth, parametric models in the regions of interest could offer a more balanced alternative.

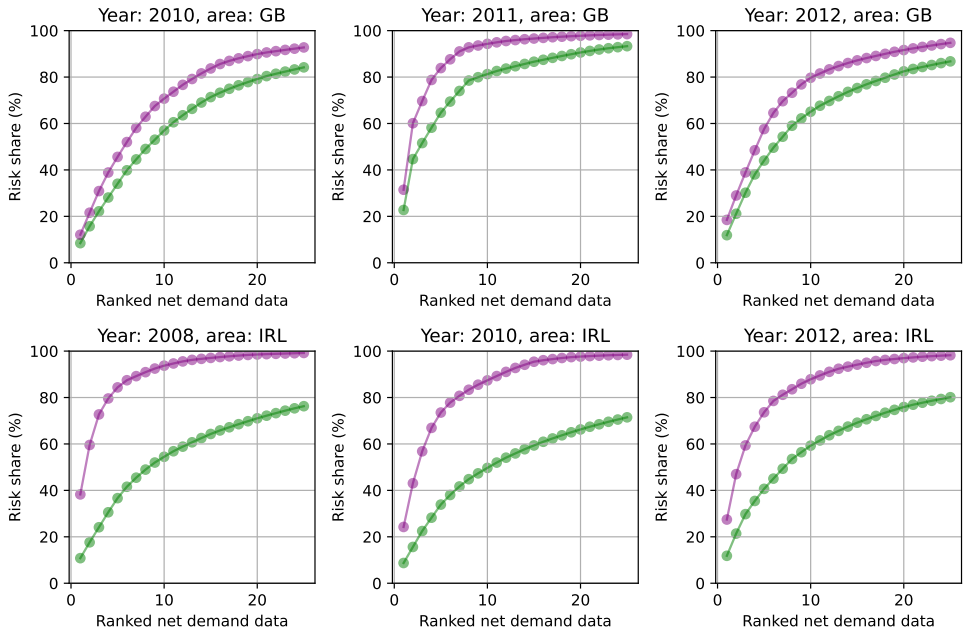


Figure 4.1: Attribution of pre-interconnection (green) and post-interconnection(purple) LOLE proportion for the highest net demand measurements in each peak season under a hindcast net demand model. For instance, just three demand measurements virtually determine post-interconnection LOLE in 2012 for IRL (lower right corner).

As the influence of net demand on shortfall risks is almost completely limited to the upper tail of net demand data in both systems, i.e., from the highest values of (demand minus available wind), this section describes the application of methodology from the theory of extremes to model net demand tail events in

the two-area system, modelling each peak season separately but using the same methodology for all of them. Then, we compare risk estimates of this model to those of an empirical (hindcast) model of net demand. All of this was done using a bespoke Python package `riskmodels` which is publicly available through the PYPI repository (Sanchez, 2022).

4.5.1 Parametric models for net demand extremes

4.5.1.1 Fitting GPD tail models

We first fit univariate Genralised Pareto distributions for each system using the largest net demand observations in the peak season under consideration; in order to set exceedance thresholds, we follow the approach described in (Coles, 2013) using mean residual life plots, concluding that 95% quantile thresholds are appropriate for both areas in all individual peak seasons; then, using the exceedances above said thresholds we fit generalised Pareto models. For both areas and all peak seasons we observe negative fitted shape parameters (ξ in (4.9)); this means net demand data has light tails in the sense that the fitted models have a finite upper endpoint. Fig. 4.2 shows Q-Q plots for the fitted tail models in both areas. We observed similar goodness of fit for net demand in all years. Finally, fitted parameters are shown in Table 4.2.

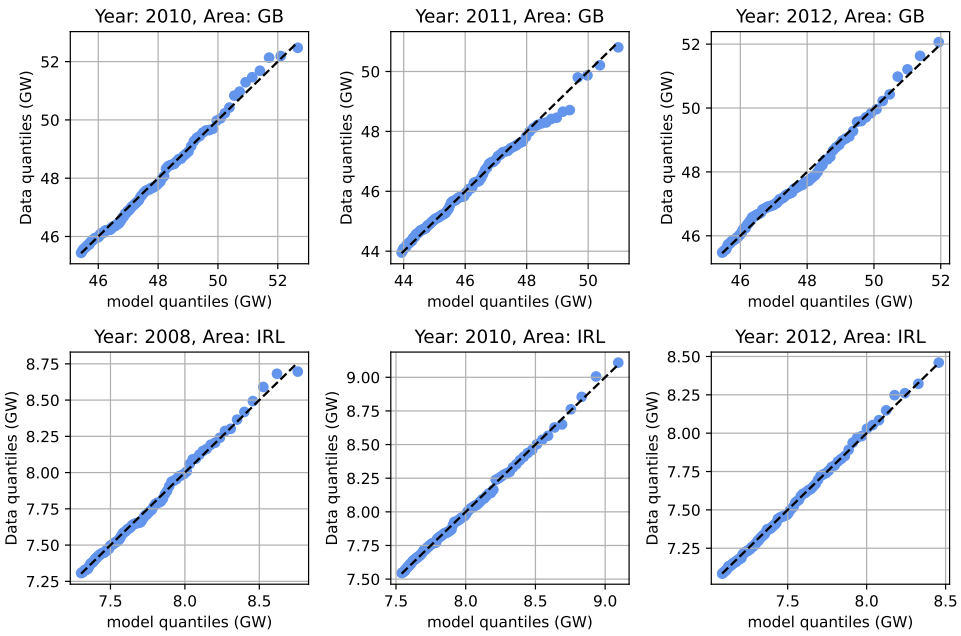


Figure 4.2: Q-Q plots for tail models in both areas. Data shown are exceedances above the 95% quantile threshold for each peak season, and the fitted models are generalised Pareto distribution.

4.5.1.2 Testing for extremal dependence in net demand

Common visual diagnostic based on empirical estimates for χ were consistent with the hypothesis of asymptotic dependence for all years except 2011 and 2012 (Figure C.1); for the latter, confidence bounds for empirical estimates of (4.11) contained 0 for some of the largest quantiles. We also attempt to obtain a numerical assessment of this hypothesis by performing a Bayesian ratio test on the coefficient of tail dependence η (Section 4.3.2) using the 95% quantile as threshold for both systems. As the hypothesis of asymptotic dependence consists of a single point in the hypothesis space, namely $\mathcal{H}_0 : \eta = 1$ vs $\mathcal{H}_1 : \eta \neq 1$, $\eta \in [0, 1]$, we calculate the Savage-Dickey ratio r, defined as the ratio between the posterior and prior densities for the value of interest (Chen, 2005; Wagenmakers et al., 2010), in this case $\eta = 1$. This can be thought of as the limiting value of a Bayes ratio test when the null hypothesis subset shrinks to a single point; the test value itself, r, equals the ratio of posterior to prior odds for $\mathcal{H}_0$ (Dickey and Lientz, 1970).

Year	ratio
2007	7.24
2008	1.6
2009	4.75
2010	2.55
2011	2.79
2012	0.15

Table 4.1: Results for the Savage-Dickey ratio test described Section 4.5.1.2 for each peak season of net demand

To perform the test, we set $\mathbb{P}(\mathcal{H}_0) = \mathbb{P}(\mathcal{H}_1) = 0.5$ and use a Jeffreys prior on both parameters η and σ; Jeffreys priors are invariant under reparametrisation (Castellanos and Cabras, 2007), which makes them a robust choice when no additional information is available. We restrict η to $[0, 1]$ and treat $\sigma > 0$ as a nuisance parameter. Under this setting, r values larger than 1 make $\mathcal{H}_0$ more credible than the alternative, and we find this to be the case for all individual years except 2012 (Table 4.1). Note that even though some values might appear to provide only weak evidence for $\mathcal{H}_0$, the fact that it is at the very edge of the hypothesis space might make the posterior converge slowly to $\mathcal{H}_0$ even when it is true, and indeed we observed similar values for r when using synthetic data for which $\mathcal{H}_0$ was the correct choice (Figure 4.3). The use of uniform priors did not alter the conclusions of the tests, and we thus conclude asymptotically dependent extreme value models are appropriate for our net demand data.

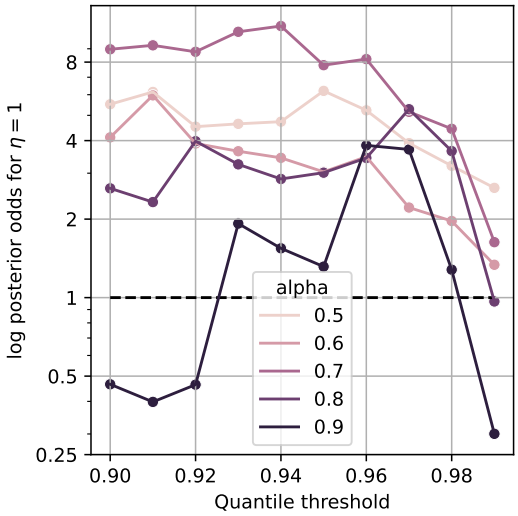


Figure 4.3: Results of the Savage-Dickey test described in Section 4.5.1.2 on a peak season's worth of synthetic data with logistic dependence (Equation 4.15). The test correctly identifies the existence of asymptotic dependence in most scenarios, but posterior odds remain relatively low; some false negatives occur for $\alpha = 0.9$, the closest tested value to asymptotic independence, i.e., $\alpha = 1$.

Lastly, as asymptotic dependence was the favored hypothesis for all other years, for the sake of simplicity we use the same model for 2012 too, rather than alternatives such as (Heffernan and Tawn, 2004). There is further discussion of the validity of this model of asymptotic dependence in Section 4.6.1 (which considers dependence between surpluses, as opposed to the net demands studied here).

4.5.1.3 Fitting the dependence model

To fit the dependence models we use the same modelling thresholds as for the marginal exceedance models, and we consider logistic and asymmetric logistic dependence (Section 4.3.3). A visual comparison of the empirical Pickands approximation induced by joint exceedance data (as calculated in (Hall and Tajvidi, 2000)) against those of the models fitted on joint exceedances (Fig. 4.4) suggests that both models provide an appropriate descriptions of the data, except for 2008 where the asymmetric logistic produces a better fit. However, the effect of the choice between these two models turns out to be minimal in the consequent results, and so we choose a logistic model for parsimony.

Lastly, the model is fitted in all of the exceedance region, i.e., that in which at least one exceedance in the two areas occur. Fitted dependence parameters are shown in Table 4.2.

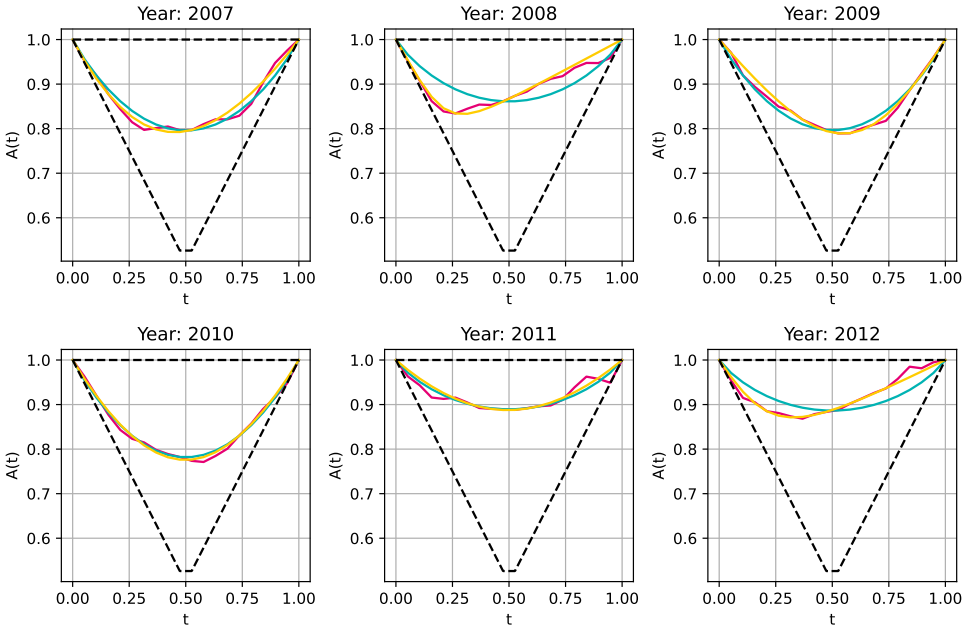


Figure 4.4: Comparison of the fitted logistic (teal) and asymmetric logistic (yellow) Pickands functions to the empirical Pickands function (red) for each peak season.

4.5.1.4 The fully fitted model

Given a sample of net demand data $\boldsymbol{y}_1, ..., \boldsymbol{y}_n \in \mathbb{R}^2$ and a quantile threshold $q \approx 1$ (corresponding to the modelling threshold μ in Section 4.3.1), the steps below summarise the fitting and sampling process for the final bivariate model used hereafter. Note that for data samples, component indices are denoted as superscripts and sample indices as subscripts:

1. Fit the semiparametric model in Eq. (4.10) for each component (i.e., for each single-area system). Let us call the CDF functions of the fitted univariate models' $\bar{F}_1(y), \bar{F}_2(y)$.

2. Project samples to copula scale $\boldsymbol{u}_1, ..., \boldsymbol{u}_n \in [\boldsymbol{0}, \boldsymbol{1}]$ where

$$\boldsymbol{u}_i = \left(\bar{F}_1\left(y_i^{(1)}\right), \bar{F}_2\left(y_i^{(2)}\right)\right)$$

   Define $\hat{\boldsymbol{U}}$ as the empirical copula of the projected sample, $\boldsymbol{q} = (q, q)$ and

$$\kappa = 1 - \mathbb{P}\left(\hat{\boldsymbol{U}} \leq \boldsymbol{q}\right) = \mathbb{P}(\hat{\boldsymbol{U}} \not\leq \boldsymbol{q})$$

3. Let $\boldsymbol{\mathcal{U}}$ be the parametric copula given by Eq. (4.16), and fit the conditional tail model $\boldsymbol{\mathcal{U}} \,|\, \boldsymbol{\mathcal{U}} \not\leq \boldsymbol{q}$ to projected data in Step 2 in the region $\boldsymbol{u} \not\leq \boldsymbol{q}$. Note

that in general the CDF of a conditional model of the form $\boldsymbol{\mathcal{U}} \,|\, \boldsymbol{\mathcal{U}} \nleq \boldsymbol{q}$ can be written as

$$F_{\boldsymbol{\mathcal{U}} \,|\, \boldsymbol{\mathcal{U}} \nleq \boldsymbol{q}}(\boldsymbol{u}) = \frac{F_{\boldsymbol{\mathcal{U}}}(\boldsymbol{u}) - F_{\boldsymbol{\mathcal{U}}}(\min\{\boldsymbol{q}, \boldsymbol{u}\})}{F_{\boldsymbol{\mathcal{U}}}(\boldsymbol{q})}, \; \boldsymbol{u} \nleq \boldsymbol{q}$$

Where $\min\{*\}$ is applied component-wise. The density of this conditional model is proportional to the unconditioned model, and so maximising likelihood is straightforward. As the fit is performed only with data in the region $\boldsymbol{u} \nleq \boldsymbol{q}$, the fitted parameters are those that better represent dependence at the tails; the fitted model only cover said region, and the case for $\boldsymbol{u} \leq \boldsymbol{q}$ is addressed below.

4. We now define the full semiparametric copula model $\bar{\boldsymbol{U}}$ along the same lines of Eq. (4.10); in this case its CDF is given by

$$F_{\bar{U}}(\boldsymbol{u}) = \begin{cases} F_{\hat{U}}(\boldsymbol{u}) & \text{if } \boldsymbol{u} \leq \boldsymbol{q} \\ F_{\hat{U}}(\min\{\boldsymbol{q}, \boldsymbol{u}\}) + \kappa \cdot F_{\boldsymbol{\mathcal{U}} \,|\, \boldsymbol{\mathcal{U}} \nleq \boldsymbol{q}}(\boldsymbol{u}) & \text{if } \boldsymbol{u} \nleq \boldsymbol{q} \end{cases}$$

5. Sample $\bar{\boldsymbol{u}}_1, ..., \bar{\boldsymbol{u}}_n$ from $\bar{\boldsymbol{U}}$ and map to original scale $\bar{\boldsymbol{y}}_1, ..., \bar{\boldsymbol{y}}_n$ by doing

$$\bar{\boldsymbol{y}}_i = \left(\bar{F}_1^{-1}\left(\bar{u}_i^{(1)}\right), \bar{F}_2^{-1}\left(\bar{u}_i^{(2)}\right) \right)$$

The components of the semiparametric distribution $\bar{\boldsymbol{U}}$ are sampled according to their probability mass, i.e., $1 - \kappa$ for $\hat{\boldsymbol{U}}$ and κ for $\boldsymbol{\mathcal{U}} \,|\, \boldsymbol{\mathcal{U}} \nleq \boldsymbol{q}$. The former is an empirical copula and is straightforward to sample. The latter is more complex, but in the case of a two-dimensional Gumbel copula corresponding to a logistic model, this can be done efficiently by using inverse transform sampling (Fishman, 1996) as follows: let $U^+ = \max\{u_1, u_2\}$ and $U^- = \min\{u_1, u_2\}$, then sample u^+ from $U^+ \,|\, U^+ > q$ and then u^- from $U^- | U^+ = u^+$; finally, one of the resulting vectors (u^+, u^-) and (u^-, u^+) is chosen at random. For the copula corresponding to (4.17) the method described can be used as a basis for sampling too, since an asymmetric logistic model can be sampled as elementwise maxima of independent logistic models (Stephenson, 2003). However, for more than 2 dimensions or different copulas, other sampling methods might be required. Once data are sampled, they can be projected back to the original data scale using the numerical inverse functions of (4.10) for each component.

In other words, the fully fitted model is semiparametric and comprises the

empirical data distribution for the non-extreme region $\boldsymbol{Y} \leq \boldsymbol{\mu}$, with threshold vector $\boldsymbol{\mu}$ given in Table 4.2, and a parametric exceedance model defined on the region $\boldsymbol{Y} \not\leq \boldsymbol{\mu}$. The model's marginal distributions are as in (4.10) with fitted tail parameters as in the referenced table, and dependence in the exceedance region is given by the fitted copula.

season	GB (μ, σ, ξ)	IRL (μ, σ, ξ)	dependence (α)
2007	(43.18, 2.78, -0.33)	(6.96, 0.37, -0.08)	0.53
2008	(44.7, 2.58, -0.31)	(7.3, 0.5, -0.21)	0.51
2009	(45.42, 2.05, -0.23)	(7.44, 0.48, -0.28)	0.47
2010	(45.4, 2.9, -0.3)	(7.54, 0.5, -0.18)	0.46
2011	(43.91, 2.66, -0.26)	(6.7, 0.52, -0.2)	0.56
2012	(45.43, 2.41, -0.26)	(7.08, 0.47, -0.21)	0.6
2013	(42.45, 2.48, -0.36)	(6.85, 0.53, -0.27)	0.5

Table 4.2: Table with fitted model parameters for all peak seasons. GB and IRL columns show the fitted parameters for the respective univariate exceedance distributions; μ, σ are in GW, and μ correspond to the 95% quantile in all cases.

4.5.2 Comparison to hindcast net demand models

In this subsection, we compare LOLE estimates from the modelling methodology described in previous subsections to those from the hindcast model (4.4).

To make this comparison more relevant in the context of growing renewables penetrations and market integration, we perform numerical experiments using an interconnection capacity of 2 GW and a wind generation capacity of 2.5 times that from installed in 2014 (i.e., installed wind capacities of 7.5 GW for IRL and 38 GW for GB). The distribution of surplus in each system is then shifted to keep the pre-interconnector value of LOLE averaged across all seven years at the initial level of 3 h/y.

Both a hindcast and a semiparametric logistic methodology are used to model net demand data in each of these scenarios, and the resulting risk estimates are compared below.

4.5.2.1 Comparison of LOLE estimates

Fig. 4.5 shows post-interconnection LOLE estimate for both areas; because the integral in (4.7) cannot be calculated exactly for the fitted tail models, Monte Carlo estimates are used instead, and corresponding confidence bands for the central estimate are shown in blue (results for all years and both areas can be

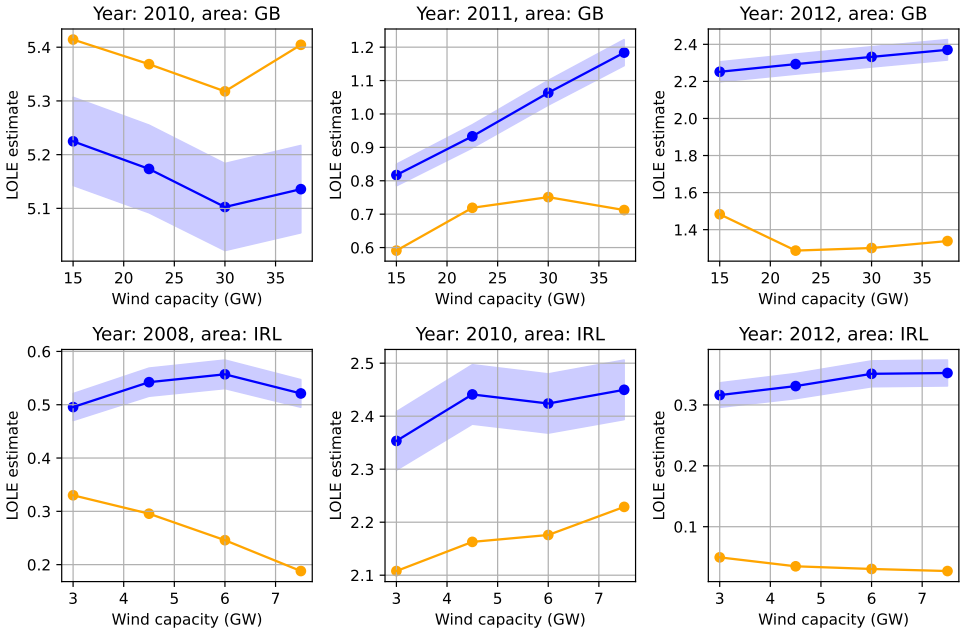


Figure 4.5: Comparison of LOLE estimates from hindcast (orange) and logistic (blue) models. Because estimation from a logistic model requires Monte Carlo simulation, 95% confidence bands for its central estimate are shown.

seen in Figures C.4 and C.5). As these results consider the contribution of interconnection, LOLE values are generally low, particularly for Ireland where the system is smaller relative to the capacity of interconnection. However we can observe large relative differences between model estimates at both areas, e.g. years 2011, 2012 for GB and 2008, 2012 for IRL. Furthermore, for 2011 in GB and 2008 in IRL, both models suggest diverging risk estimates as more wind is installed. While the difference between the hindcast and EVT results are not very great for the highest risk year (2010), there are substantial differences in some of the higher risk years (2012 in GB, 2008 in Ireland), demonstrating how the effect of smoothing the tail can be material in practical risk calculations.

Lastly, the next Section contains a discussion on different attempts to compare both net demand models in terms of output quality and estimation variability. However, results were inconclusive, and it was not clear whether a fair comparison was even possible in this regard, as it would most likely need a credible generative probability model for net demand traces, which if existed would obviate the need for either of the tested models in adequacy calculations. Conversely, although fair comparisons between both models can be made on synthetic data sets, it is not clear how to evaluate whether these data sets are representative of net demand data in an appropriate sense; however, the discussion is included for the sake of completeness.

4.5.2.2 On the comparison of estimation variability between hindcast and EVT-based net demand models

Figure 4.5 shows that both models can produce visibly different risk estimates in some scenarios, as well as different trends on the evolution of LOLE as more wind capacity is installed. Investigating which model offers more accurate risk estimates is thus a valid undertaking. We already saw that hindcast LOLE estimates tend to be based on very few historic observations (Figure 4.1), which would point to high output variability. Parametric models of extremal dependence can help on smoothing out this coarseness, but we have also seen that determining the existence of asymptotic dependence (which enables the justified use of said models) is difficult and there is a risk of dependence type misspecification, which would produce an overestimation or underestimation of risk; this is in addition of possible model misspecification errors within the correct type of extremal dependence.

Making a valid comparison between the models is not straightforward, as it is not possible to obtain different realisations of a given set of winter conditions with which this can be done directly. In this section we attempt to assess model output quality by using synthetic net demand data sampled from an appropriate 'ground truth' net demand distribution, repeatedly computing outputs from both models using these synthetic peak seasons and assessing their output mean and confidence bounds around it; we compare these outputs to the 'ground truth' risk, e.g. the risk induced by the ground truth net demand distribution. Thus, for the following experiments we select generative 'ground truth' distributions from which 500 realisations of synthetic peak seasons are simulated, all of which are used to produce output LOLE values from both models. In the case of logistic semiparametric models, they are fitted from scratch for each simulated peak season following the methodology outlined in previous sections. However, we found that devising an appropriate ground net demand distribution to make a valid comparison turned out to be more difficult than anticipated; below we describe the approaches we tried.

We first use the fitted semiparametric net demand model from the previous section as ground truth model. We note that this is a time-collapsed model, i.e., it produces i.i.d samples and as such it is not representative of real net demand data, which exhibits a high degree of autocorrelation. However, it can be argued that this model preserves the most important features to reproduce the behaviour of net demand data in a relevant sense, e.g. tail behaviour and extremal dependence. Asymptotic guarantees from EVT would arguably also make this choice relatively

uncontroversial regarding marginal distributions. The lack of autocorrelation structure would likely make estimation variability appear artificially small, but it is likely that this phenomenon is of similar magnitude across both models, thus i.i.d data would still be valid to explore relative differences between models.

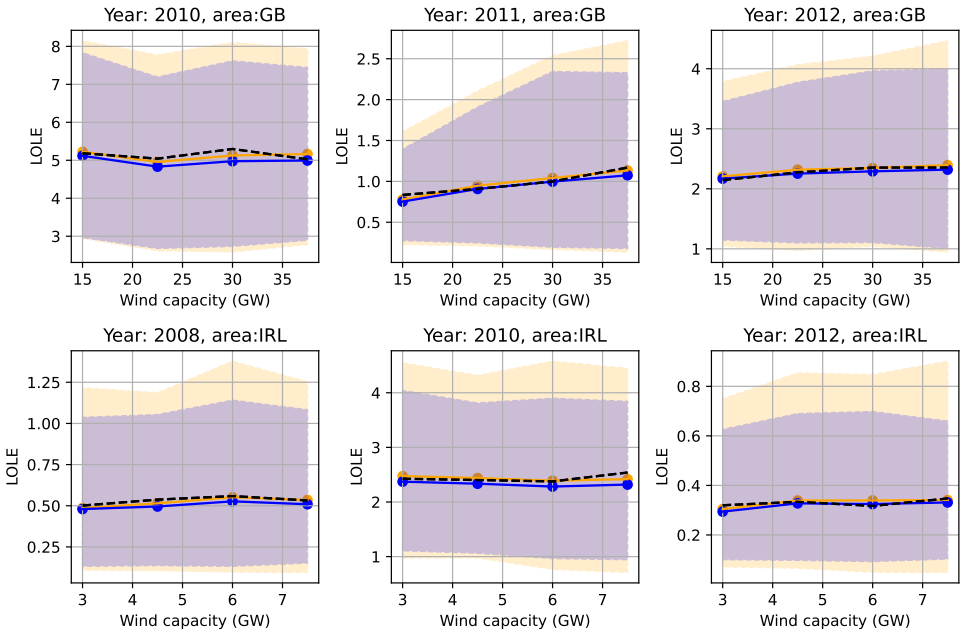
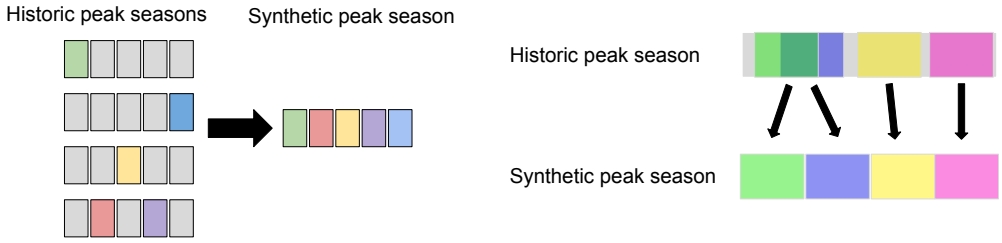


Figure 4.6: Comparison of sampled post-interconnection LOLE estimates for hindcast (orange) and logistic (blue) models, for sampling distributions based on different years in the historic record. Solid lines show average estimated LOLE, while filled sections show 95% confidence bands. While both models recover the true sampling distribution's LOLE (black dotted line) on average, hindcast models show consistently higher estimation variability.

Figure 4.6 shows the results for this experiment based on different years for the ground truth model; 95% confidence bounds are shown (results for all years and both areas can be seen in Figures 4.9 and 4.10), representing the variability across the 500 samples drawn. Taking these bounds as a measure of the corresponding estimation variability, we observe higher hindcast-based variability across all years and scenarios. Moreover, the relative difference for some years is considerable, particularly for Ireland, whose size is smaller relative to interconnection capacity. For instance, results for 2012 in IRL show 40% more variable hindcast estimates compared to a logistic model, with a considerably higher upward bias; this was also true for 2007, which is not shown here.

There is a significant problem with the experiment above, however, which is implicitly assuming there is no model misspecification whatsoever, i.e., the ground truth model belongs to the same parametric family than one of the models be-



(a) First approach. Synthetic peak seasons are assembled by a blocked sampling approach from multiple historic peak seasons

(b) Second approach. Single peak seasons are recombined through week-long block sampling. Sampled blocks can overlap.

Figure 4.7: Diagrams showing how blocked bootstrap sampling was performed in experiments from section 4.5.2.2

ing tested, which could make the comparison unfair, even if a hindcast model is entirely non-parametric. For this reason, our second experiment tries instead a blocked bootstrap approach using historic net demand data. Each historic peak season in our data consists of 20 consecutive weeks. With this in mind, two approaches were tried: in the first one, synthetic peak seasons were randomly assembled by choosing each of its 20 week blocks uniformly at random from the corresponding blocks of historic peak seasons (Figure 4.7a). Nevertheless, this incurs in issues related to mixing winter conditions from different years, complicating the interpretation of any takeaways from this experiment. For instance, we have seen in Figure 4.4 and Table 4.1 that different peak seasons likely exhibit slightly different extremal dependence structures. Thus, the mix of different peak season data would likely result in a mixture dependence model unrelated to any relevant physical process. The second approach, on the other hand, performs week-long blocked resampling from within a given peak season, uniformly sampling 20 one-week periods (not necessarily aligned to calendar weeks; see Figure 4.7b).

The second approach avoids the mentioned issues from mixing different winter conditions, and emulates the within-peak-season autocorrelation structure to a certain degree. Results for this experiment using 500 synthetic peak seasons are shown in figure 4.8 (results for all years and both areas can be seen in Figures 4.11 and 4.12). In this case we observe significantly more variability from logistic models than for hindcast models. However, direct comparison of confidence bounds might not be entirely appropriate due to the fact that there are also significant

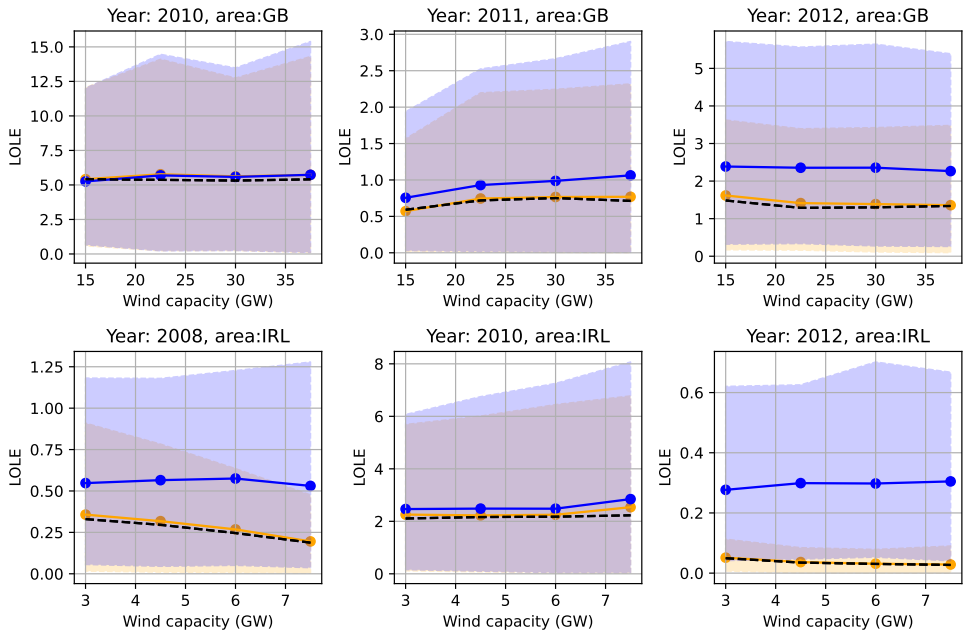


Figure 4.8: Comparison along the same lines of figure 4.6 for a blocked bootstrap resampling of each of the displayed peak seasons. In blue, results for logistic model outputs; in yellow, results for hindcast model outputs. Dotted black line shows estimated risk from ground truth model; shaded regions denote 95% confidence bands.

differences in the models' mean risk estimates (in other words, a standard error of, say, 0.5 denotes a different degree of uncertainty for a mean value of 3 than for a mean value of 0.6). In fact, mean estimates in said figure for both models are exactly those from Figure 4.5. this suggests that this experiment may now be biased towards the hindcast model, as the ground truth distribution reproduces its results exactly. This makes sense, as it is essentially re-sampling historical observations, but again complicates any conclusion on comparing the quality of outputs from both models.

From the last two experiments, it looks like designing a credible, unbiased ground truth net demand distribution would require significantly more work than any of the simple models tried above. We explored using linear and non-linear models to de-trend net demand, using the resulting mean trends together with randomly sampled residuals to produce synthetic peak seasons. However, diagnostic plots revealed that a good deal of residual autocorrelation persisted in both cases. It is possible that autocorrelation has some impact on risk (as in reality autocorrelation would be driven partly by weather patterns such as cold spells). Moreover, post-interconnection LOLE from these models were somewhat different than those from original data and so it was not entirely clear to which

degree these models were a valid representation of net demand data.

Despite not being able to get conclusive results on comparing model output quality, we believe EVT models offer a more mathematically principled approach to risk management than hindcast models, and provide aditional functionality such as calculation of probabilities beyond the data range. Even for peak seasons in which it is not sufficiently clear whether a model for asymptotic dependence is the most appropriate one, outputs from an EV-based model like the one developed in this section could be used as a credible upper bound on post-interconnection risk.

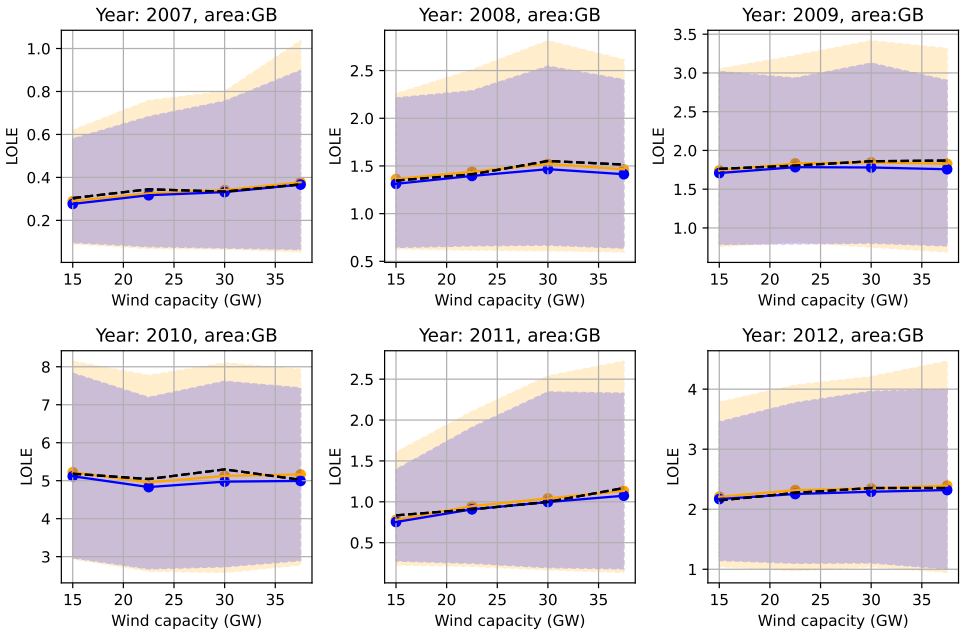


Figure 4.9: Model output comparison (all years) for GB in the logistic ground truth model experiment as outlined in Section 4.5.2.2.

4.6 Statistical dependence in capacity surpluses

Having developed a model for extreme net demand events in the two-area system in the previous section, we now turn our attention to statistical dependence in capacity surpluses, i.e., between the values of (available supply minus demand) in the two areas. The analysis in the previous section demonstrated a strong association in extreme *net demand* co-occurrences, however each of these net demands is convolved with the available conventional capacity in the relevant area, and the available conventional capacities are assumed independent. It is thus natural to think that in some relevant sense the dependence may be weaker

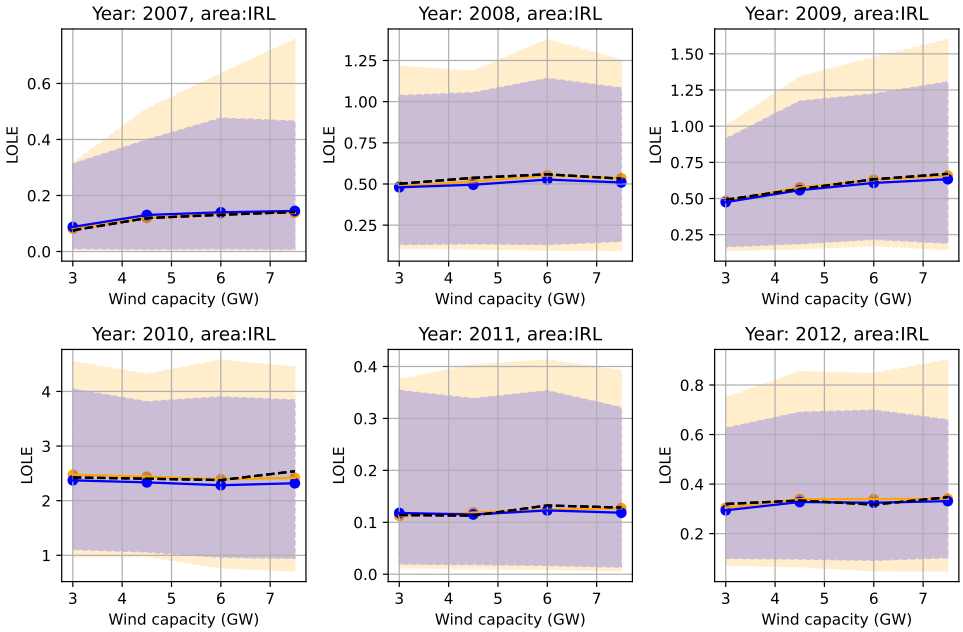


Figure 4.10: Model output comparison (all years) for IRL in the logistic ground trugh model experiment as outlined in Section 4.5.2.2.

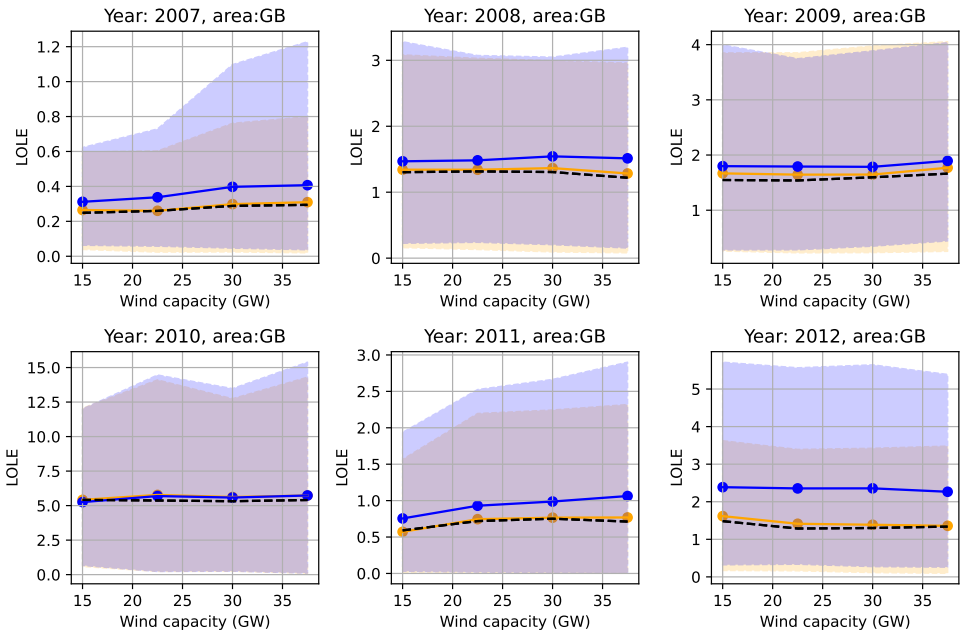


Figure 4.11: Model output comparison (all years) for GB in the blocked bootstrap experiment as outlined in Section 4.5.2.2.

for power *surpluses* than for *net demand*.

We work with the negative of capacity surpluses, treating shortfalls as maxima instead of minima to make results from EVT immediately applicable. To avoid any confusion we call this the *shortfall distribution*, and negative values simply

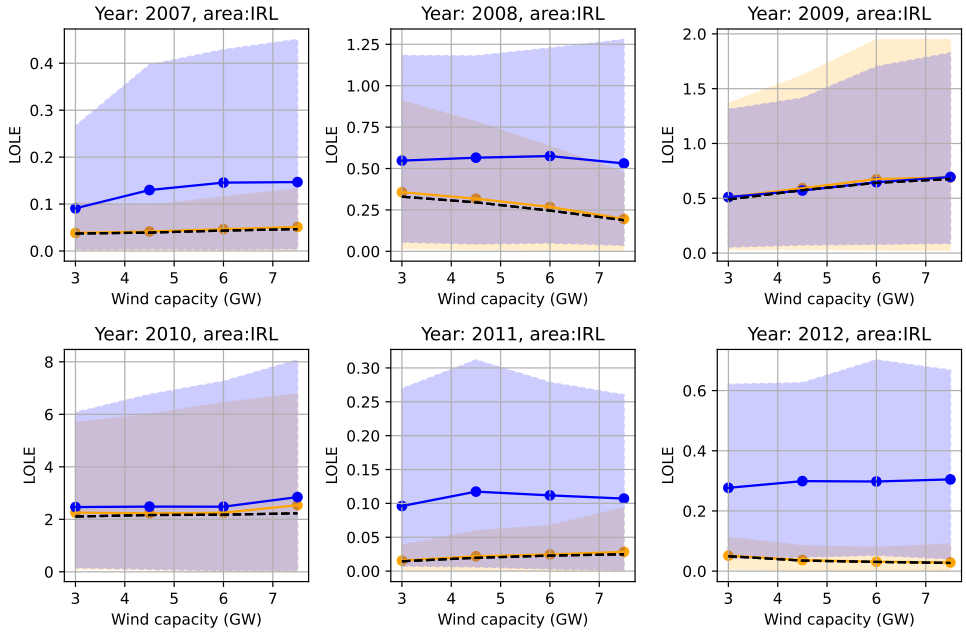


Figure 4.12: Model output comparison (all years) for IRL in the blocked bootstrap experiment as outlined in Section 4.5.2.2.

indicate the non-occurrence of a shortfall.

4.6.1 Characterising statistical dependence

We proceed in a manner similar to the analysis of net demand in Section 4.5, performing the same Bayesian ratio test as in Section 4.5.1.2 to determine whether asymptotic dependence is present, resulting in a Savage-Dickey ratio of $r < 10^{-3}$ for all years and a wide range of quantile thresholds between 80% and 99.99%. This provides strong direct evidence of capacity shortfalls being asymptotically independent across areas.

The evidence for asymptotic dependence between the two net demands was not as definitive as this evidence against for capacity shortfalls, but the difference in r for the two cases clearly demonstrates stronger tail dependence for the net demands as compared to the surpluses. This would appear to confirm the intuition that the convolution with the (independent) conventional plant distributions should weaken the dependence between the systems – and provides some further justification for the use of the logistic model for the earlier cases even where the evidence for asymptotic dependence was relatively weak.

Extreme value copulas like the ones used on net demand data are not appropriate to model asymptotically independent data. However, as was said in Section 4.3.2, Gaussian copulas provide a simple parametric example of an asymptotically

independent copula, and we test this as a dependence model for joint exceedances of the shortfall distribution, finding that it is a good fit for all peak seasons and tested thresholds between 80% and 99.99%. Fig. 4.13 shows the comparison between contour lines for each decile of the empirical and fitted Gaussian copula for a threshold of 80%.

We note that Gaussian copulas are not typically used to model extremal dependence in the EVT literature because they are not an extreme value copula in the sense of Equation (4.13). Our choice here is instead based on practicality, as it provides a simple parametric model that accurately describes tail dependence in the capacity shortfall distribution.

Based on this evidence, we proceed as with the net demand model and use a Gaussian copula to describe dependence between exceedances at both components, defining an exceedance as a shortfall in the corresponding component. The modelling region is illustrated in Fig. 4.14b. The quantile threshold in this case was much higher than for net demand, as shortfalls occur with a probability of approximately 0.001% in each area, due to the LOLE normalisation to three hours per year over the whole seven-year period.

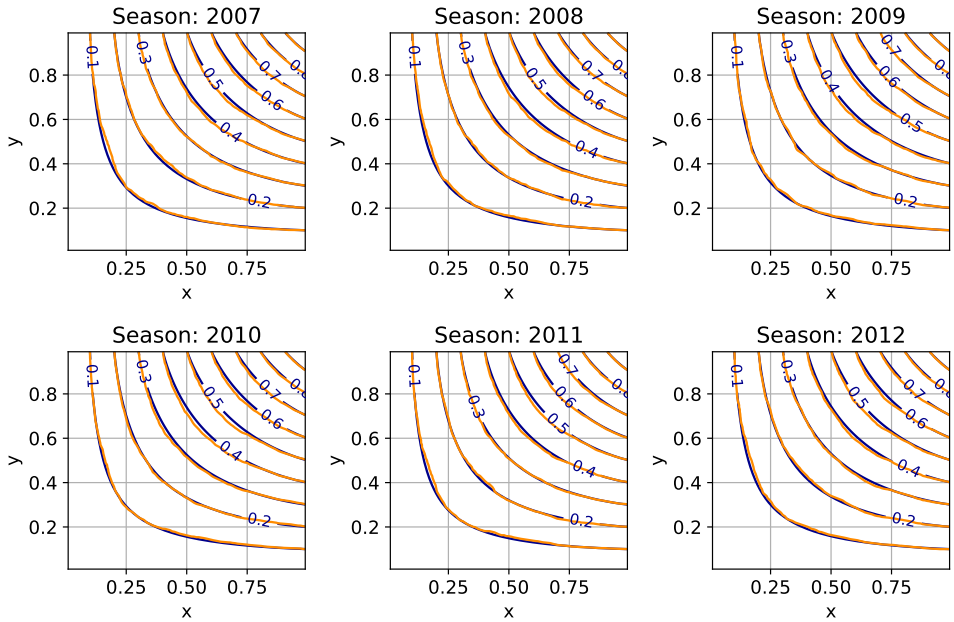
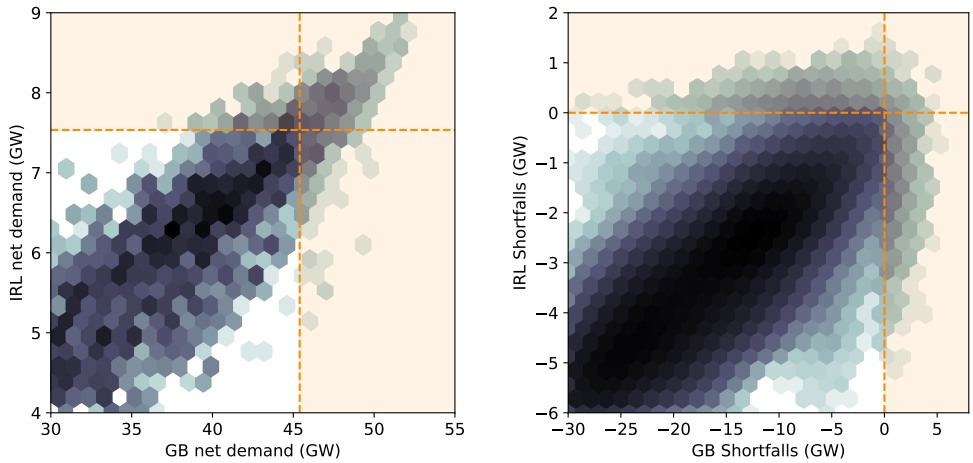


Figure 4.13: Contour lines for cumulative probability functions of empirical (orange) and fitted Gaussian copulas (blue); the first nine deciles are shown. The data consisted of 2500 simulated joint shortfalls.

The difference between the regimes of tail dependence in net demand and the shortfall distributions can be more clearly seen in Fig. 4.14: the net demand distribution's tails are much more concentrated around the diagonal of the graph



(a) Net demand data

(b) Simulated shortfalls

Figure 4.14: Hex-binned net demand and capacity shortfall distribution scatterplots for 2010. The modelling thresholds, which delimit the modelling region (shaded orange) are shown as dashed lines. Points in the modelling regions were simulated from the fitted models.

than those of capacity shortfalls. A consequence of this is that simultaneous shortfalls occur rarely relative to shortfalls in just one region, whereas there is a much stronger tendency for extremes of net demand to occur in both areas at the same time. Fitted dependence parameters for this model are given in Table 4.3.

season	ρ
2007	0.68
2008	0.7
2009	0.69
2010	0.76
2011	0.7
2012	0.66

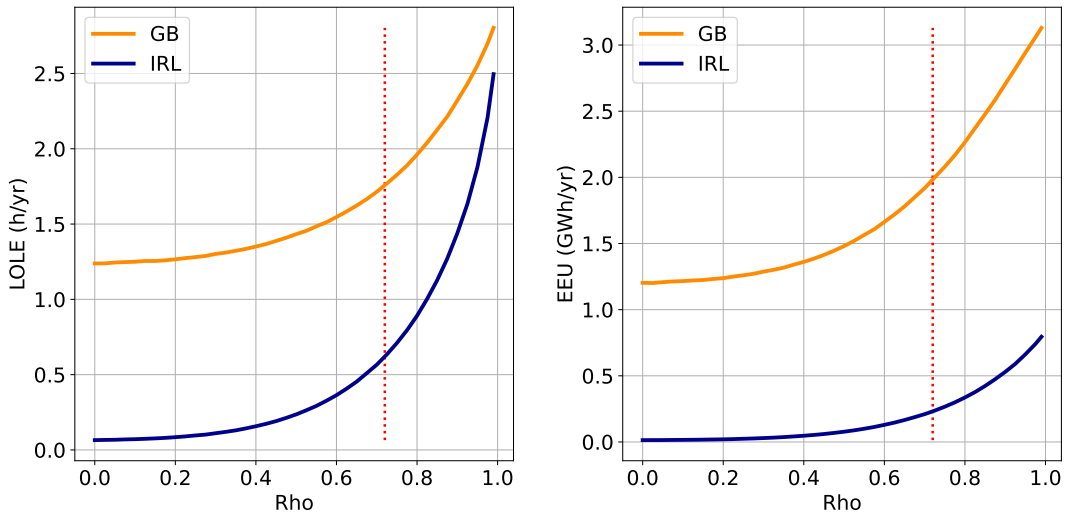
Table 4.3: Fitted values for Gaussian exceedance dependence model.

We note that these observations about dependence between deficits in the two areas follows in substantial part from the assumption of independence between available conventional capacities in the two areas. This assumption is usually made in practical calculations, though it must be caveated if there is a possibility of a common cause event affecting units in both systems – for instance restriction on primary fuel supply, or elevated failure rates in certain weather conditions as in (Murphy et al., 2018, 2020). Nevertheless, these results around asymptotic dependence are helpful in understanding outputs of present risk calculations, and it seems reasonable to think that even where common cause events are relevant

the influence of conventional plant will weaken the tail dependence between the distributions of deficit in the two areas as compared to the dependence between net demands.

4.6.2 Sensitivity of risk metrics to dependence strength

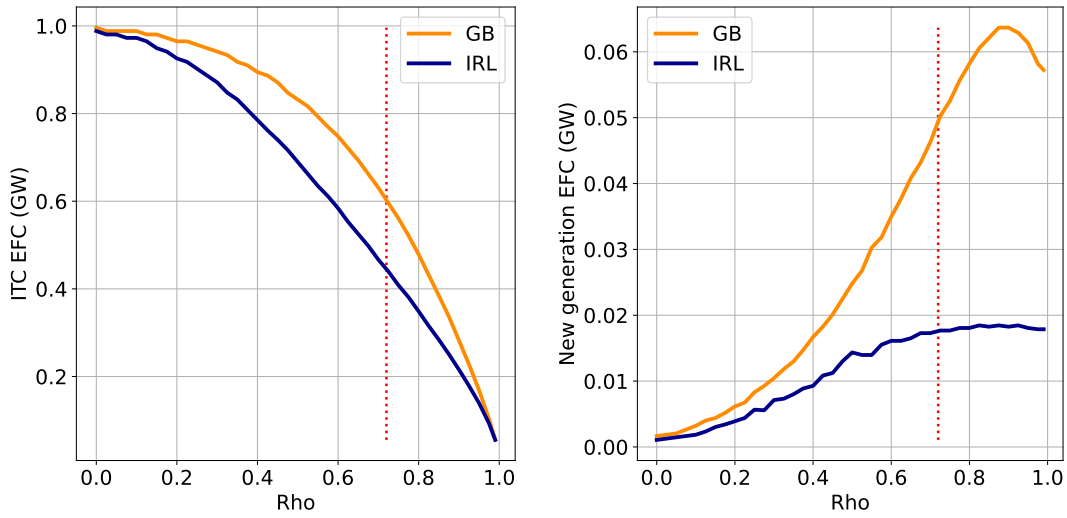
Having a single-parameter dependence model for the shortfall region in the capacity shortfall distributions makes it straightforward to perform a sensitivity analysis of LOLE and EEU to statistical dependence strength between the capacity surpluses at both areas. Fig. 4.15 shows the results of the sensitivity analysis using all available historic data. The dependence of the LOLE level is quite weak at low values of ρ, with the estimated risk level only increasing significantly above the 'independence' limiting case forρ above about 0.5; as the latter approaches 1, LOLE values approach pre-interconnection levels.



(a) LOLE (b) EEU

Figure 4.15: Sensitivity of each metric (LOLE and EEU) to dependence strength, in both areas; the dotted line represents the estimated dependence parameter in the GB-IRL system.

Lastly, we also perform a sensitivity analysis for the equivalent firm capacity (EFC) of interconnection and of new non-local conventional generating units that are installed on the other end of the interconnector, looking at how their EFC depends on the strength of statistical dependence between capacity deficits across areas and using LOLE as the risk metric with which this is computed (see (Zachary and Dent, 2012) for a more detailed overview of capacity value metrics). While for the area where new generation is being installed, a new unit's



(a) Interconnection's EFC (b) New generating unit's EFC

Figure 4.16: Sensitivity of EFC for interconnection and new generating units installed outside each system (i.e., on the other side of the interconnector); the dotted line represents the estimated dependence parameter in the GB-IRL system.

EFC equals its expected available capacity at a randomly chosen point in time when measured through LOLE, for the area at the opposite side this depends on multiple factors such as each system's installed capacity background, their shortfall-sharing policy and interconnection capacity. Moreover, this can be useful to quantify from the point of view of capacity market operations, e.g. if generating units from other systems were allowed to participate in local capacity auctions. To demonstrate this we assume a nameplate capacity of 100 MW and an availability probability of 95% for the new unit. The results of the analysis are shown in Figure 4.16b.

We can see in Figure 4.16a that for systems with statistically independent capacity deficits, interconnection's EFC equals 1 GW, i.e., the assumed interconnection capacity in these experiments. This is a consequence of the fact that under this scenario, available imports can saturate the interconnector whenever one of the areas experience a capacity deficits, as simultaneous occurrences of low capacities are extremely unlikely. As statistical dependence grows, the EFC is reduced at a roughly constant rate, until being close to zero for $\rho = 1$. It is interesting to note that for a range of values roughly around 0.5 for the dependence parameter ρ, interconnection's EFC is visibly larger for GB than for IRL; this could be a reflection of the fact that capacity shortfalls in IRL on average require less assistance from GB than vice versa simply because of the difference in system size.

It is to be expected that in Figure 4.16b the EFC of the new non-local generating unit (e.g. at the other side of the interconnector) initially increases for both areas as statistical dependence become stronger. This is because systems with statistically independent capacity surplus/deficits would be very likely to have enough spare capacity to completely mitigate a capacity shortfall in the other area or saturate the interconnector, following the same argument as above. Thus, for statistically independent systems new non-local capacity would make little difference to the already considerably reduced post-interconnection risks. As dependence strength increases, this start to change, and additional non-local generation start to become more valuable.

The peak and subsequent decrease in EFC for new Irish generation from the point of view of GB is likely due to the fact that in our experiments IRL has a slightly larger LOLP than GB; this means that as $\rho \to 1$ and capacity deficits at both systems start to become perfectly dependent, the Irish system would start experiencing shortfalls whenever the British system does, but not the other way around, and consequently newly installed Irish capacity would have a progressively smaller impact on GB under a veto policy, since any new unit would be completely used up by Irish demand at times when GB might benefit from them; hence, $\rho = 1$ would imply EFC $= 0$ from the point of view of GB (note the plotted values only reach $\rho = 0.99$).

The EFC for new generating units in GB from the point of view of IRL has a milder slope than vice versa in Figure 4.16b. This is likely a consequence of the difference in system size. The GB system is already large enough to substantially reduce shortfalls risks in IRL by means of exported capacity, so new generation in GB will have a smaller effect in IRL than the other way around. As $\rho \to 1$, the argument from the previous paragraph would also explain the sudden increase in EFC for new units in GB.

4.7 Conclusion

This Chapter has presented approaches to statistical modelling of net demands and of capacity shortfalls in two area power systems. The general approach is to assess whether the quantities of interest are asymptotically dependent, and then to choose an appropriate bivariate model based on this assessment.

In the case of net demands, the tests are consistent with asymptotic dependence, though this conclusion is not very definitive. Other diagnostics do also suggest that an extreme value copula with asymptotic dependence provides a

good fit to the empirical dependence structure. The resulting smoothing of the sparse extreme region of the empirical distribution of net demand can change risk calculation results substantially, with a possible explanation being that the calculation results involving the empirical distribution are driven by a very small number of historic records.

We also demonstrate how the dependence structure of capacity shortfalls in the two areas may be modelled directly. In this case there is very strong evidence against asymptotic dependence, and we find that a Gaussian copula describes this behaviour appropriately in this case. This contrast in the assessment of asymptotic dependence for the two cases confirms the intuition that dependence between shortfalls in the two systems might be weakened by convolution with the distributions of available conventional capacity (which are assumed independent between the systems). This also provides a means of performing sensitivity analysis of the risk model outputs to the strength of dependence between the shortfalls, by varying the correlation coefficient of the Gaussian copula.

4.8 Discussion

The net demand model developed in this Chapter shows high goodness of fit for the GB-IRL system; moreover, from a statistical point of view, models based on EVT offer more principled tools for adequacy calculations than hindcast models, and offer other advantages that we did not fully explore in this Chapter, such as the ability of calculating exceedance probabilities beyond the historic data range, or using the estimated parameter values on relevant sensitivity analysis calculations, such as measuring dependence strength under different wind penetration scenarios as we did in Chapter 3. We also argue that EVT models as used here provide the right level of granularity and appropriate mathematical tools for the problem of adequacy assessment, particularly when compared to market simulation modelling tools that use much more granular models (down to hourly unit commitment simulations) but are not designed for this problem. More specifically, such a high level of granularity might ultimately compound uncertainty in the estimations, which are not designed to model the rare, extreme events that drive adequacy risks.

Lastly, as we are working here on a time-collapsed framework, the models we have developed can only really be used to compute time-collapsed risk metrics such as LOLE and EEU or perform sensitivity analysis of quantities based on these risk indices, but the risk profile of a power system is complex and cannot

be summarised by these metrics alone: in (Sheehy et al., 2016), an example was shown of important changes in the system's risk profile that are not accounted for by LOLE as more wind capacity is installed into the system, namely an increase in the probability of severe loss of load events. It is reasonable to think that other important properties of the system could also change which are not reflected by these indices; the progression to an increased reliance on wind capacity currently taking place in GB and other systems is a structural transformation that potentially alters the distribution of loss of load events in complex ways. It is critical to understand these changes in order to integrate renewable technologies without disrupting electricity supply. To look at statistics beyond LOLE, however, a sequential model of the system is needed. Sequential models also require the use Monte Carlo estimation for risk indices as the methodology in this Section, but they are more computationally demanding. In the next Chapter, we explore large scale sequential modelling to look at more detailed features of the system's risk profile and explore issues that may be relevant to decision-makers beyond LOLE or EEU risk indices.

Chapter 5

An analysis of the impact of very high wind capacities on system adequacy beyond long-term averages

So far in this thesis, we have worked on a time-collapsed framework, using standard risk indices like loss of load expectation (LOLE) and expected energy unserved (EEU) to characterise the system's risk under different scenarios, and have developed statistical methodology on top of this picture. These indices have long been used to asses system adequacy both by researchers and system operators as we saw in Section 2.2. However, these are just long-term averages of the underlying quantities of interests, and their usage alone might offer too narrow a view of the system's risk profile for some of the practical considerations with which system planners are faced: for instance, questions about variability of outcome of the corresponding quantities of interest, or the relative frequency of events of a certain magnitude cannot be addressed through these indices. It is reasonable to believe that for most decision-makers, a system with low variability would be preferable to one with high variability even if both have the same risk in terms of LOLE or EEU. Decision-makers might thus exhibit some degree of risk-aversion (Grechuk and Zabarankin, 2014; Kahneman and Tversky, 1979), which raises the need to explore system risk beyond long-term averages. This is particularly relevant under the current transition to a high wind capacity regime in Great Britain (GB), which could alter the statistical behaviour of loss of load events in complex ways, many of which may be of considerable importance for policy makers yet are not reflected by the usual risk indices.

In this Chapter, we use a sequential model representative of the (single-area) GB system to look at the system's risk profile beyond long-term averages such as LOLE and EEU, and illustrate concrete examples of the issues discussed above as more wind is installed: through extensive simulation, we explore the probability distributions associated to different aspects of loss of load events such as their duration, size and spread across peak season days, looking at how these and other quantities of interests change in the context of an increasing reliance on wind generation. We focus this analysis on the evolution of the system through the coming decade, using data on projected installed wind capacity in GB up to the year 2032. Moreover, we also explore the impact of this technological transformation on the capacity procurement decision process in GB and incorporate a notion of risk aversion into the current formulation. Our results suggest that as the installed wind capacity reach the projected levels for 2032 of approximately 71 GW, the resulting system has substantially higher variability of outcome than in low wind regimes, with fewer loss of load events of relatively mild severity (as expressed by the quantities of interest mentioned above) but considerably more events of high severity. This progression occurs even if wind capacity is installed in such a way that LOLE or EEU indices do not change. Moreover, there is also a progression toward a system with fewer days of potentially much more severe shortfalls than a system with low or no wind capacity. Even more specifically, one where hourly loss of load events tend to increasingly occur grouped together into fewer but longer runs, defining a run as a sequence of one or more consecutive hourly time indices. The overall increase in variability also impacts the decision process for capacity procurement, where optimal procured capacity becomes more sensitive to parameters such as the value of lost load (VOLL) and the degree of risk aversion from the decision-maker as the system's wind capacity grows.

5.1 Introduction

Reliability of electricity supply is a crucial consideration in energy system planning. In some systems, like the one in Great Britain (GB), capacity markets have been set up to ensure there is enough investment in new capacity to satisfy reliability standards (DECC, 2013c). When a decision on the amount of capacity to procure is required, this is usually based either on target values for reliability standards, or on a cost-benefit analysis which considers the costs of procuring new capacity and the costs associated to the system's reliability level, the latter of which is usually calculated using the value of lost load (VOLL) parameter, which

represents the average costs to consumers per unit of energy not served (National Grid, 2017b). In either case, decisions are often based on a standard set of risk indices such as loss of load expectation (LOLE), expected energy unserved (EEU), or other summary statistics such as system average interruption frequency index (SAIFI) or the system average interruption duration index (SAIDI) (Billinton and Allan, 1985).

However, there are multiple potential issues with this picture from the point of view of decision analysis. Firstly, as pointed out above, it is reasonable to believe that decision-makers dealing with rare but potentially very negative outcomes as is the case with capacity shortfalls might prefer to avoid risk, i.e., they might exhibit some degree of risk aversion. Risk aversion is a central concept in decision theory, and has been widely studied (Meyer, 2014). It can be understood as the decision-maker's disinclination to gamble (Smith, 2010). More specifically, a risk-averse agent will always prefer a deterministic pay-off of x to a random pay-off with an expectation of x. Under the usual framework of expected utility theory, this property is a consequence of the agent's utility function being concave, which is related to the concept of diminishing returns (Grechuk and Zabarankin, 2014). A risk-averse decision-maker would thus be interested in aspects beyond long-term averages, such as variability of outturn for individual years. This is by definition not possible to express by expectation-based risk indices, e.g. LOLE and EEU. Thus, reliability standards based on long-term averages alone might not provide a good reflection of the concerns of decision-makers.

Secondly, expressing reliability standards as a single numerical value as is usually done through LOLE or EEU might provide too narrow a view of the system. Multiple characteristics of loss of load events are relevant for system planners, including duration, frequency and the amount of lost load, among other statistics. Hence, system adequacy exhibits a multidimensional risk profile that may not be appropriately summarised by any given univariate metric. This issue may become more pressing in the current context of technological change in which the share of intermittent energy sources in the generation mix is quickly growing: in (Sheehy et al., 2016), a model representative of GB was used to show that an increasing share of wind generation causes greater variability of outturn in both energy unserved and the duration of loss of load events, even if wind is incorporated in such a way that the system's LOLE remains constant. Thus, aspects that would likely be deemed highly relevant in the evolution of the system are simply not represented by the usual univariate reliability standards. Issues with trying to summarise system risk in a single number have been noted before: in (Tindemans and Strbac, 2015), this is addressed by proposing more comprehensive

multi-dimensional risk profiles assembled from the distribution of outputs from a sequential simulations of the system. A clustering procedure is then used to summarise the outputs into a limited number of qualitatively different scenarios that can be evaluated by decision-makers.

Thirdly, the estimation of the VOLL parameter in cost-benefit analyses for capacity procurement as currently done in GB can be highly problematic: there is ample disagreement in how to measure it, even when only considering the impact of loss of load events on aspects for which a monetary value results natural, such as economic losses (Schröder and Kuckshinrichs, 2015; Billinton et al., 1983). Moreover, it is recognised that loss of load events can have an impact on less tangible aspects that are difficult to value monetarily even though they ultimately have an economic effect, such as a decrease in wider economic confidence or foreign investment (National Grid, 2017b). Furthermore, it is usually assumed that during a loss of load event, the system operator will be able to disconnect the precise amount of load required and nothing more, but this might not be the case in practice, which would effectively make VOLL larger than assumed, possibly substantially. For instance, in the case of GB it is noted in (Ofgem, 2013) that the number and type of disconnected customers (e.g. commercial, residential or industrial) cannot be known in advance and depends on the composition of demand at the time of loss of load. Thus, it is reasonable to believe that at the time of a capacity shortfall there would be considerable uncertainty as to how to proceed with disconnections in order to minimise societal or economic costs. Lastly, even if one accepts the idea of monetising energy unserved in the usual way, it is not clear that the average value per customer is an appropriate summary, given that effects on different customers can vary widely, and so a larger value more representative of customers for which this is a major difficulty might also be worth considering.

In this Chapter, we aim to show concrete examples of the issues outlined above by performing an analysis of a model representative of the GB system, focusing the analysis on aspects that might be considered important by decision-makers.

We do this in two ways: firstly, we perform large scale sequential simulation of the system's capacity surplus/deficit, looking at how the progression to a very high wind regime (with 71 GW of installed wind capacity) alters the probability distributions associated to different aspects of loss of load events, including their duration and energy unserved, as well as their spread across the peak season. We assume the systems meets current reliability standards throughout this progression, and in this sense, we work on a similar picture to (Sheehy et al., 2016) but perform more extensive simulations with a larger set of updated future wind

scenarios, and looking at a wider set of loss of load statistics; we also test the sensitivity of our results to different experimental configurations.

Secondly, we incorporate a measure of the decision-maker's degree of risk aversion into the cost-benefit analysis framework currently used in GB for capacity procurement. Under the current methodology, costs associated to the system's reliability level are calculated by multiplying the EEU by an estimate of the VOLL to compute the total unreliability costs. We replace the EEU by a conditional value at risk (CVaR) (Embrechts et al., 1999; Artzner et al., 1999) of energy unserved. CVaR can be calculated as an expectation conditional on an exceedance above a user-defined quantile level; in this sense, it generalises the EEU. Moreover, the chosen quantile level can be seen as a measure of risk aversion. We use the resulting cost function to explore how the optimal capacity to procure changes as a function of the decision-maker's degree of risk aversion and the VOLL parameter as more wind is installed into the system.

This Chapter is organised as follows: Section 5.2 outlines the current picture of adequacy assessment in GB; Section 5.3 reviews the current methodology for capacity procurement in GB; Section 5.4 provides an overview of sequential Monte Carlo modelling methodology for adequacy assessment; Section 5.5 outlines the data we use to perform the experiments; Section 5.6 and 5.7 describes the experiments and the results we obtained, and finally, Section 5.8 discusses the findings of this Chapter and possible extensions.

5.2 Current reliability standards in GB

Recall the formulation of the adequacy assessment problem that we have used throughout this work, given by Equation (2.3), and let us assume without loss of generality that the period of interest (usually a peak season) is divided into n hourly segments. Current reliability standards in GB are based on having a LOLE of 3 hours/year, and the estimation methodology relies on the use of time-collapsed models for available conventional capacity (ACC) and net demand (defined as demand minus wind generation), the latter of which is forward-mapped to the period of interest (Ofgem, 2013; NGESO, 2021c). Sections 2.1 to 2.5 cover modelling methodology for demand, wind and ACC in more detail.

The resulting time-collapsed model of net demand, Y, is given by the empirical distribution of the forward-mapped historic data $y_1, ..., y_\eta$, and can be written as

$$\mathbb{P}(Y \leq y) = \frac{1}{\eta} \sum_{\tau=1}^{\eta} \mathbb{I}(y_\tau < y) \tag{5.1}$$

where τ is the time index from the historic period and $\mathbb{I}(\cdot)$ denotes an indicator function. A time-collapsed probability model for ACC, X, can be calculated numerically using data on the number and type of conventional generating units as well as their historic availability rates. The assumption of independence between Y and X produces the following identities for LOLE and EEU

$$\text{LOLE} = n \cdot \mathbb{P}(X < Y) = n \cdot \frac{1}{\eta} \sum_{\tau=1}^{\eta} \mathbb{P}(X < y_\tau) \tag{5.2}$$

$$\text{EEU} = n \cdot \mathbb{E}[(Y - X)_+] = n \cdot \frac{1}{\eta} \sum_{\tau=1}^{\eta} \mathbb{E}_X[(y_\tau - X)_+] \tag{5.3}$$

where $(*)_+ = \max\{*, 0\}$.

5.3 The capacity procurement problem in GB

After assessing the risk for the scenario under consideration, decision-makers are tasked with determining whether investment on any additional capacity is needed, and if so, how much of it to procure. In the case of GB, this is done through the capacity market, in which auctions are held every year to procure capacity for 4 years into the future (Engie, 2016). The decision on how much capacity to procure through these auctions corresponds to the Secretary of State, but their decision is supported by a formal recommendation from the system operator (Ofgem, 2013). This recommendation is based on a cost-benefit analysis methodology that aims to balance economic costs from procuring additional capacity with the costs associated to the resulting reliability levels.

Approaches based on a cost-benefit analysis are part of what is sometimes called value based reliability assessment (Billinton and Allan, 1985), and in general consist in procuring an amount of capacity $x^* \in \mathbb{R}$ such that it minimises the total economic costs $c(x)$, given by

$$c(x) = \kappa(x) + r(x) \tag{5.4}$$

where $\kappa(x)$ is the capacity procurement cost, which is an increasing function of the procured capacity x, and $r(x)$ is a cost associated to the reliability level (or lack thereof) after procuring x units of capacity, and is a decreasing function of x; in the rest of this Chapter we refer to $\kappa(x)$ and $r(x)$ as *procurement costs* and *unreliability costs*, respectively.

Elicitation of $c(x)$ under the current methodology used by National Grid

makes the assumption that procurement costs are proportional to the amount of capacity procured; additionally, the unreliability costs are taken to be proportional to the amount of expected energy unserved. This can be formalised by the following cost function

$$c^{\mathrm{GB}}(x) = \beta \cdot x + \gamma \cdot \mathrm{EEU}(x),\ \beta, \gamma > 0 \tag{5.5}$$

$$\mathrm{EEU}(x) = \mathbb{E}\left[\sum_{t=1}^{n} (Y_t - X_t - x)_+ \cdot \Delta_t\right] \tag{5.6}$$

Note that calculation of the unreliability costs in (5.6) assumes procured capacity will only shift the distribution of the capacity surplus/deficit process. This is a reasonable assumption when the amount of procured capacity is small with respect to what is already installed into the system, and does not include new generation that may be statistically associated with existing generation, e.g. wind turbines (Dent and Zachary, 2014). Both of these conditions are often met in capacity market auctions. Note also that since EEU is a time-collapsed quantity, (5.3) can also be used to compute $\mathrm{EEU}(x)$ in the obvious way.

In practical calculations, values for β and γ in (5.5) are given by estimates of the cost of new entry (CONE) and value of lost load (VOLL) respectively. The former is taken to be approximately £47,000/MW-year, and is based on the estimated annual revenue required to cover the construction costs of a peaking open cycle gas turbine (OCGT) (DECC, 2013a,b,c,d). VOLL is taken to be approximately £17,000/MWh, and is based on the estimated average value to consumers of preventing disconnections at times of peak demand (DECC, 2013c,d); note both of these values are annualised and expressed per unit of energy, i.e., MWh in the case of VOLL and £/MW/y in the case of CONE. The cost function used in GB for annual recommendations on capacity procurement thus takes the form

$$c^{\mathrm{GB}}(x) = \mathrm{CONE} \cdot x + \mathrm{VOLL} \cdot \mathrm{EEU}(x) \tag{5.7}$$

Due to the high uncertainty of what will be connected to the system multiple years into the future, National Grid contemplates a set of different scenarios when producing its recommendation. This set of scenarios $\mathcal{S}$ considers different system configurations and sensitivities in both the supply and demand side, which are informed by market intelligence and consultation with subject matter experts (NGESO, 2021c). Each scenario $i \in \mathcal{S}$ defines a cost function $c_i^{\mathrm{GB}}(x)$ of the same form as (5.7), which incorporates the corresponding scenario assumptions through

the distributions of X_t, Y_t in (5.6); for convenience, let us denote the associated optimal decision as $x_i^*,\ i \in \mathcal{S}$.

Once the set $\mathcal{S}$ is elicitated, a least worst regret (LWR) decision rule (Loomes and Sugden, 1982; Bell, 1982; Fishburn, 1982) is used to make the final recommendation on the amount of capacity to procure, x^*_{LWR}, given the set $\mathcal{S}$ and its associated cost functions $\{c_i^{\text{GB}}(x)\}_{i\in\mathcal{S}}$. LWR is closely related to the min-max decision rule (Savage, 1951), for which the output x^*_{minmax} is the decision that minimises the maximum cost across scenarios; for the problem at hand this is given by

$$x^*_{\text{minmax}} = \min_{x\in\mathbb{R}}\{\max_{i\in\mathcal{S}}\{c_i^{\text{GB}}(x)\}\}$$

The calculation of x^*_{LWR} requires to first define a *regret function* for each scenario $i \in \mathcal{S}$, where we take regret to be the difference between the optimal cost and the decision's cost. These functions can be written as

$$c_i^{\text{LWR}}(x) = c_i^{\text{GB}}(x) - c_i^{\text{GB}}(x_i^*),\ i \in \mathcal{S}$$

We thus have that for all $i \in \mathcal{S}$

$$\begin{aligned} c_i^{\text{LWR}}(x) &\geq 0\ \forall x \in \mathbb{R} \\ c_i^{\text{LWR}}(x_i^*) &= 0 \end{aligned}$$

This is, for each scenario regret is always non-negative, and equals zero only if the optimal decision for the corresponding scenario is chosen. The LWR solution minimises the largest regret across scenarios, and for the problem at hand is given by

$$x^*_{\text{LWR}} = \min_{x\in\mathbb{R}}\{\max_{i\in\mathcal{S}}\{c_i^{\text{LWR}}(x)\}\}$$

Note that this is equivalent to the min-max solution on the regret functions.

5.4 Time-sequential adequacy models

LOLE and EEU form the basis of reliability standards in GB and many other power systems (National Grid, 2017b). These indices are the expected values of two common quantities of interest in adequacy calculations, namely the loss of load duration (LOLD), defined as the number of time units (in this case, hours) in which a shortfall will have occured, and the energy unserved (EU), defined as the

overall amount of power that will fail to be supplied. These are mathematically defined as

$$\text{LOLD} = \sum_{t=1}^{n} \mathbb{I}(X_t < Y_t) \cdot \Delta_t \tag{5.8}$$

$$\text{EU} = \sum_{t=1}^{n} (Y_t - X_t)_+ \cdot \Delta_t \tag{5.9}$$

where Δ_t is the time step length (usually expressed in hours), Y_t, X_t are net demand and ACC at time t respectively, and the sum is taken over the future period of interest. However, statistics other than their expected value (say, their quantiles or standard deviation) are not time-collapsed in nature, i.e., there is no analogous to (5.2) and (5.3) and their calculation requires the use time-sequential models and Monte Carlo estimation.

There is abundant literature on the use of Monte Carlo methods for applications in adequacy assessment where exact estimation is not possible (Li, 1994). The main challenge when using sequential Monte Carlo methods is their high computational cost. LOLD, EU and related quantities are determined by rare events, as was discussed in Chapter 4. For this reason a large number of realisations might be required for any estimate to attain a reasonable level of accuracy. This problem has been addressed in various ways in the literature, for example, (Billinton and Jonnavithula, 1997) explores variance reduction techniques in the context of adequacy calculations; (Tindemans and Strbac, 2020) explores the use of multi-level Monte Carlo (MLMC), in which estimates from a sequence of models with increasing degrees of granularity are combined to speed up computations while preserving accuracy. More recently, MLMC methods have also been combined with machine learning based surrogate models to speed up the elicitation of the multiple submodels involved (Sharifnia and Tindemans, 2022).

Sequential hindcast models for net demand simply assume a repeat of historic net demand conditions, as the time series of forward-mapped historic data are used directly (Sheehy et al., 2016). In the case of ACC, time-sequential models like those from Section 2.3 are typically used, where individual generating units are modelled as (usually two-state) Markov chain processes. In this case, in addition to historic failure rates, data of mean times to repair (MTTR) or mean times to failure (MTTF) from generating unit types are used to estimate the chain's transition probabilities (Singh et al., 2018; Billinton and Allan, 1985). The aggregated availability X_t is then given by a sum of independent Markov

chain processes. Under this model, Monte Carlo realisations of LOLD and EU conditioned to a peak season $\mathcal{S}$ in the forward-mapped data are given by

$$\text{lold} = \sum_{t \in \mathcal{S}} \mathbb{I}(x_t < y_t) \cdot \delta_t \tag{5.10}$$

$$\text{eu} = \sum_{t \in \mathcal{S}} (y_t - x_t)_+ \cdot \delta_t \tag{5.11}$$

where y_t are the forward-mapped net demand values corresponding to historic peak season $\mathcal{S}$, and x_t are simulated values for ACC. Statistics from these and other random variables can then be estimated in the usual way conditioned to particular peak seasons or for the entire period. Additionally, calculation of other relevant quantities beyond LOLD and EU from the resulting realisations, e.g. number of days with capacity shortfalls, is also straightforward.

5.5 Data

We use openly available historic demand data from the GB system for the period 2005-2017 (NGESO, 2022). Demand met by generation embedded at the distribution level was added back to demand metered at the transmission level in order to obtain the system's total demand[1]. Furthermore, we use peak season dates and Average Cold Spell (ACS) peak factors from Table D.1 to rescale historic peak season demand data to a common baseline of 55,000 MW. The latter value is only meant to be representative of demand levels in GB (see Section 2.4 for details on the ACS peak factor rescaling methodology). In total, the demand data used here comprises 12 peak seasons, which in GB occur during winter, from 2005/2006 to 2016/2017. From here on, we refer to individual peak seasons by the year in which they started. e.g. the peak season of 2005 corresponds to 2005/2006.

For wind generation data, we use hourly wind capacity factors for the same periods as in the demand data; these capacity factors have been reconstructed from calibrated MERRA2 reanalysis wind speed data as done in (Staffell and Pfenninger, 2016). Briefly, global reanalysis weather models such as MERRA2 provide estimates of meteorological features on an evenly spaced grid on the Earth's surface and atmosphere and at a certain time resolution, usually hourly. The resulting wind speed data can be calibrated and interpolated to wind farm

[1] this follows a similar argument than in the ACS peak factor calculation methodology in GB; see Section 2.4

locations, then combined with data on the farm's number and type of turbines to reconstruct the farm's hourly capacity factors, this is, their generation output expressed as a proportion of the farm's total installed capacity. These factors for individual farms are then combined to produce capacity factors at the national level; see Section 2.5 for more details.

Wind data sets resulting from the process described above for GB are also openly available (Pfenninger and Staffell, 2015), and comprise both offshore and onshore capacity factors for three different wind fleets: the *current* scenario corresponding to wind farms that were operational as of 2016, the *near term* scenario corresponding to wind farms that were in some construction stage as of 2016, and the *long term* scenario corresponding to wind farms that were in some pre-approval stage as of 2016. We use capacity factors corresponding to the aggregated *current* and *near term* fleets, but do not include the *long term* fleet due to the high uncertainty inherent to the completion of construction projects in pre-approval stages. The resulting capacity factors can then be rescaled using data on total installed onshore and offshore wind capacity (more on this below) to obtain reconstructed historic wind generation traces.

Data on historic and projected total installed wind capacity in GB was compiled from various sources for the period 2016-2032 . For historic data, we used the UK Government's data of total installed capacity for the years of 2016-2022 (UK Government, 2022), while for projected installed capacity values we use the Future Energy Scenarios (FES) 2021 report by National Grid (NGESO, 2021a,b). This report contains a five-years forecast of total installed wind capacity (currently up to 2027), as well as four main future scenarios up to 2050, from which we take the median total installed capacity for years beyond 2027. All wind capacities as used here include both centralised and decentralised capacity (i.e., wind generation at the both the generation and distribution levels). This results in installed capacity values shown in table 5.1.

Data on the number and availability rate of conventional generating units was originally provided by National Grid. Due to the sensitive nature of this data, random noise was added to availability rates. We take the resulting data to be broadly representative of the GB system. As we will be working with time-sequential models, we use MTTR values from the 1996 IEEE Reliability Test System (Grigg et al., 1999) for each unit type to fully specify two-state Markov chain models for the generating units (see Section 2.3 for more details).

In order to test the robustness of the results to the input data for conventional generation, we also use parameters from (Murphy et al., 2018), which derives capacity-weighted MTTR and MTTF values for each generator type from an

Year	Offshore (GW)	Onshore (GW)	Total (GW)
2016	5.3	10.8	16.10
2017	7.0	12.6	19.60
2018	8.2	13.5	21.70
2019	9.9	14.0	23.90
2020	10.5	14.1	24.60
2021	13.2	14.5	27.70
2022	14.8	15.0	29.80
2023	15.1	15.5	30.60
2024	17.3	16.0	33.30
2025	21.8	16.4	38.20
2026	22.5	18.2	40.70
2027	25.3	19.3	44.60
2028	27.1	21.9	49.00
2029	31.0	23.4	54.40
2030	40.5	24.8	65.35
2031	41.4	25.7	67.15
2032	44.5	26.5	71.00

Table 5.1: Assembled data set of installed wind capacities between 2016-2032

analysis of the GADS data set (NERC, 2019). The latter comprises historic operational data for more than 8,000 generating units in the US. This provides an additional set of credible availability and MTTR parameters.

5.6 Exploring system risk beyond long-term averages

5.6.1 Experimental configurations

The main objective of this section is to illustrate how aspects of the system's risk profile that are relevant for decision-makers can experience substantial changes as more wind capacity is installed, even if LOLE and EEU indices do not change. To devise multiple scenarios of progressively higher installed wind capacity, we take an evenly spaced partition of the period 2016-2032 inclusive, and use Table 5.1 to produce the corresponding installed wind capacity values by linear interpolation for both onshore and offshore fleets. Additionally, we include a baseline scenario of no installed wind capacity for comparison. In total, 10 wind capacity scenarios are considered, which are listed in Table D.3. For each of these, historic wind traces can be computed by rescaling historic wind capacity factors by the corresponding scenario's onshore and offshore capacity values.

We use a sequential available conventional capacity (ACC) model that assumes two-state Markov chain models for generating units (see Section 2.3 for details). For each of the tested wind capacity scenarios (Table D.3), we rescale the ACC distribution via a multiplicative factor to enforce a constraint of having a LOLE of 3 hrs/year over the entire 12-year period, effectively assuming that the system meets current reliability standards in GB.

In previous Chapters, the above has been done through a change in the location of the ACC distribution, because the main focus were risk indices based on long-term averages. In this Chapter, however, we are looking at risk indices beyond their first statistical moment, and for this reason we have used a multiplicative normalisation factor on the ACC distribution, as this is arguably a better approximation to the change in ACC variability as generators are decomissioned. Even though it may also be argued this normalisation method risks underestimating the variability of the future ACC distribution, and hence estimated risk indices, we found that multiplicative and additive normalisations on the ACC distribution produced virtually identical outcomes in our experiments. Since an additive normalisation provides an upper bound in terms of future ACC variability, we can conclude the effect of this issue in the present analysis is negligible.

Note that, while we have used a change in location of the ACC distribution to normalise system risk in previous chapters, a multiplicative rescaling factor makes more sense here as it accounts for the change in variability of ACC as more generators are decommissioned, and this factor may have an effect beyond first order moments of the risk distributions, which are the focus of this Chapter.

The presented results were generated by simulation of the system's capacity surplus/deficit in Equation (2.3) on 100,000 realisations of the historic period under consideration (i.e., 1.2 million peak seasons were simulated overall, as historic demand and wind data comprise 12 peak seasons). This was done by simulating ACC traces and subtracting historic net demand traces thereafter to compute capacity surplus/deficit values.

All of this was done efficiently through parallelised Python and C implementations which are publicly available through Python's `riskmodels` package (Sanchez, 2022) (see Chapter A), and can be installed in the usual way. Overall, results were obtained in around half an hour using 5 physical cores in a system with an Intel i7-10875H CPU and a solid state drive (SSD)[2].

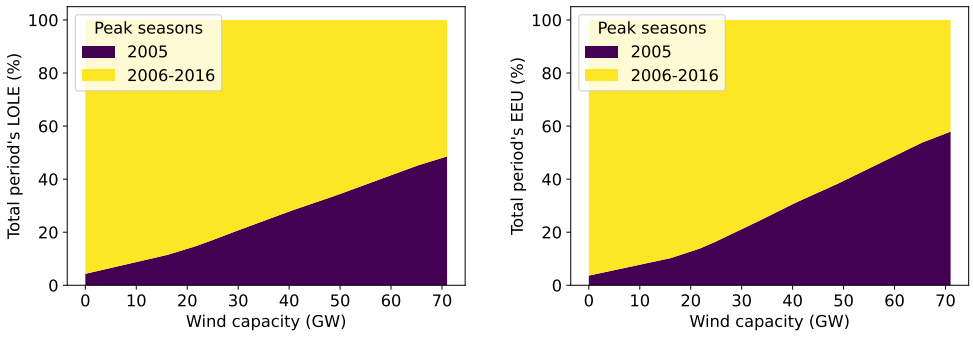
To ensure the robustness of the results presented below, we test different configurations for the system's model. The reference configuration is normalised

[2]Map-reduce operation are I/O-bound and depend heavily on storage access performance; SSDs are much more efficient than hard drives in this sense.

to have a LOLE of 3 hours/year over the period 2005-2016, and uses MTTR data from (Grigg et al., 1999) and availability data provided by National Grid. We call this the standard configuration; other model configurations are listed in Table 5.2. Results shown in the following section are for the standard configuration unless specified otherwise, and we comment on results for other configurations when relevant.

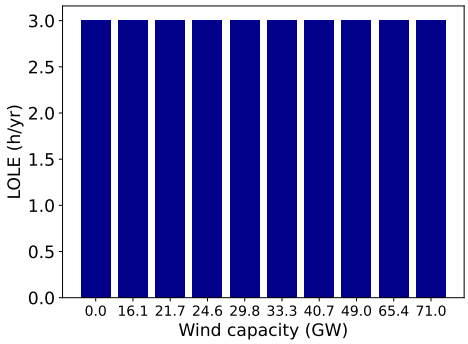
Configuration	Reliability standard	MTTR/MTTF data	period
Standard	LOLE = 3 hrs/yr	(Grigg et al., 1999)	2005-2016
low LOLE	LOLE = 0.3 hrs/yr	(Grigg et al., 1999)	2005-2016
fixed EEU	EEU = 3 GWh/year	(Grigg et al., 1999)	2005-2016
GADS parameters	LOLE = 3 hrs/yr	(Murphy et al., 2018)[3]	2005-2016
No 2005	LOLE = 3 hrs/yr	(Grigg et al., 1999)	2006-2016

Table 5.2: Table of tested experimental configurations.

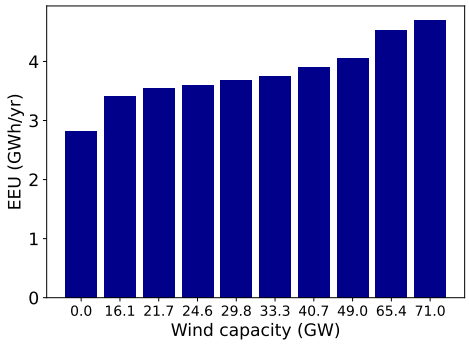


(a) total period's LOLE (%)

(b) total period's EEU (%)



(c) system's LOLE (hrs/year)



(d) system's EEU (GWh/year)

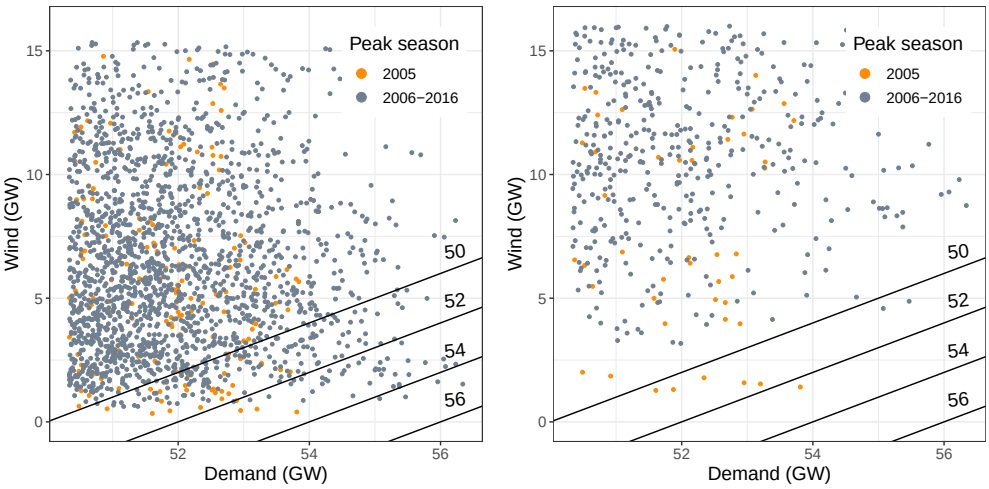
Figure 5.1: Upper row: Attribution of total LOLE and EEU in the entire 2005-2016 period for the peak season of 2005; this season concentrates an increasing proportion of the period's risk as more wind is installed. Lower row: system's LOLE and EEU for each wind capacity scenario.

[3]In this case, availability rates were also computed from the referenced article.

5.6.2 Results

5.6.2.1 Variability in year-to-year outturn becomes extreme

Some of the problems with using long-term averages to describe the system's risk in high wind regimes are apparent from Figure 5.1, which shows that the system's LOLE and EEU are increasingly determined by a single high-risk peak season in the data as more wind capacity is installed, namely that of 2005. For a wind capacity of 71 GW, the other 11 peak seasons taken together represent just half of the total risk as measured by these indices. This also means that for very high wind regimes such as these, long-term averages are not representative of typical year-to-year outturn, as the difference between the largest and smallest values of LOLE conditioned to individual peak seasons vary by many orders of magnitude: while LOLE conditioned to 2005 is 17 hrs/year, the LOLE conditioned to 2013 is just 0.08 hrs/year. Note that while LOLE is fixed across wind capacity scenarios to 3 hrs/year, the EEU consistently grows as more wind is installed (Figure 5.1d) in addition to being increasingly driven by the 2005 season. This mean loss of load events become larger on average as wind capacity grows; moreover, since the EEU is also increasingly determined by a single peak season, the extreme difference between outturn for individual years that we observed for LOLE also occurs for EEU, with conditional values for 2005 and 2013 being 32 GWh/year and 0.06 GWh/year, respectively.



(a) Wind capacity: 17 GW (b) Wind capacity: 71 GW

Figure 5.2: Comparison of wind and demand values that cause the highest observed net demand in low and high wind regimes (left and right respectively). Black diagonal lines mark contour lines for net demand.

We can more clearly see the underlying reasons for this extreme concentration

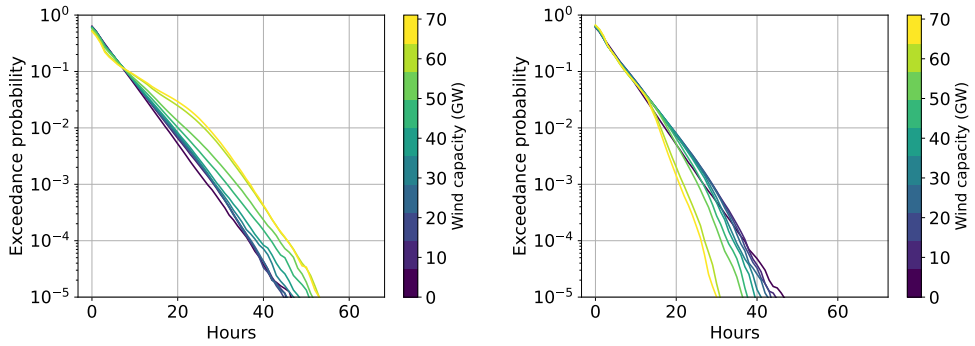
of risk in Figure 5.2. Under a relatively low wind regime of 17 GW of installed capacity, the highest net demand observations (say, in excess of 50 GW) originate from a combination of high demand and low wind (Figure 5.2a) and are spread across multiple peak seasons. However, in a high wind regime with 71 GW of installed wind capacity, it is clear from Figure 5.2b that the highest observed net demand values are characterised by low wind generation, with demand levels playing only a smaller role (as the largest observed demand levels correspond to low overall net demand levels). Moreover, the lowest observed wind generation consistently occurred in the peak season of 2005; this also suggests adequacy risks become more strongly associated with low wind generation rather than high demand in very high wind regimes.

5.6.2.2 Sampling uncertainty in the results of adequacy calculations increases substantially

In addition to the above, the inclusion (exclusion) of the 2005 peak season into the analysis determines whether extreme LOLD occurrences become more (less) frequent for very high wind capacities in the experiment's results. Figure 5.3 shows the LOLD survival functions when repeating the experiment with and without the 2005 peak season. Under 'usual' conditions (i.e., without accounting for 2005), scenarios with very high wind capacity seem to leave the frequency of events of the type $0 \leq \text{LOLD} < 15$ more or less unaffected, and reduce the frequency of more severe events (e.g. $\text{LOLD} > 20$). However, the inclusion of the 2005 peak season in the analysis changes this picture completely, with additional wind capacity consistently increasing the frequency of $\text{LOLD} > 10$ events while slightly reducing the frequency of events below this threshold. The latter trade-off is driven by normalising the system to an LOLE of 3 hrs/year, which acts by reducing the number of years with capacity shortfalls (i.e., lower frequency of $\text{LOLD} > 0$ events), as well as making occurrences of $0 < \text{LOLD} \leq 10$ less frequent, but this happens at the expense of a considerable increase in the frequency of $\text{LOLD} > 10$ events. For instance, the probability of very severe $\text{LOLD} > 25$ occurrences goes from around 10^{-3} in a baseline of no wind capacity to more than 10^{-2} for a scenario with 71 GW of wind capacity. All of this would suggest that for very high wind regimes the number of years with capacity shortfalls is lower, but when they do occur, they would potentially be much more severe than in low wind regimes as measured by LOLD. This illustrates quite clearly the general point that using long-term averages alone as reliability targets in a system that is progressing to a very high wind regime may overlook critical

considerations for decision-makers, such as a substantial increase in the likelihood of severe loss of load events (Figure 5.3a).

We emphasise the importance of the above results in a more general sense, namely, the large effect of a small number of rare, high-impact years (in this case, one) on the results and interpretation of adequacy calculations for high wind regimes: the inclusion of the 2005 peak season completely changes the results in Figure 5.3b in our experiments and the interpretation on the effect of very high wind capacities on the system's LOLD. One clear lesson from this is that the small number of historic years that are usually available for these calculations cannot accurately represent the wide range of possible weather conditions that the system can experience, some of which might entail a much larger risk than anything in the historic record for systems highly reliant on intermittent generation. Hence, the interpretation of results from adequacy studies in this case becomes highly contingent on the available historic data, which necessarily offers an incomplete picture of possible weather conditions, and it is very possible that a more complete picture in this sense could considerably alter the results, analogously to how data from 2005 alters Figure 5.3b. This issue introduces a substantial amount of uncertainty into the interpretation of these calculations, and illustrate the need to develop appropriate ways of accounting for it.



(a) Including 2005 (b) Not including 2005

Figure 5.3: Survival function for LOLD (in log scale) for two experiments with and without the peak season of 2005, which concentrates most of the period's LOLE. Each experiment is normalised to have an average LOLE of 3 hours per year for the corresponding period.

5.6.2.3 Severe energy unserved events become more frequent

Results for the distribution of EU were much more consistent across experimental configurations from Table 5.2, in the sense that in all of them the EU showed

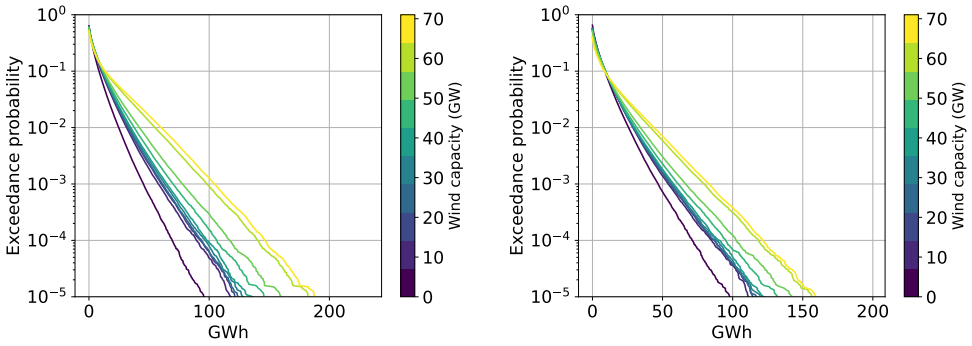
substantially heavier tails as more wind capacity is installed. This again means severe EU events become more frequent; this is consistent with the progressive increase in EEU even when keeping the LOLE constant (Figure 5.1d) as more wind is installed, nevertheless, Figure 5.4 shows that heavier EU tails also occur when looking at a system with fixed EEU across wind capacity scenarios, which suggests results on the increased frequency of severe EU events are, to an extent, unrelated to the long-term average outturn, and instead seem to be an unavoidable consequence of a increasing reliance on wind, at least in the model we have use here.

Moreover, we can also see in Figure 5.4b that as more wind capacity is installed, a reliability standard based on EEU results in a trade-off between more years without capacity shortfalls (i.e., less EU > 0 events) and a lower number of $0 <$ EU ≤ 10 GWh events at the expense of a considerable increase in the frequency of EU > 10 GWh events. For instance, the one-in-a-hundred-years return level goes from roughly 25 GWh to more than 50 GWh, and an increment of similar proportions is true for all return levels above 20 years. This is analogous to the effect on the LOLD distribution under a reliability standard based on LOLE = 3 hrs/yr (Figure 5.3a). However, we note that the increase in the frequency of severe EU events in a system that uses EEU as a reliability standard (Figure 5.4b) is less pronounced than in a system with a reliability standard based on LOLE (Figure 5.4a).

Overall, our results are consistent with those obtained in (Sheehy et al., 2016) in the sense that we observe a substantial increase in the likelihood of severe EU and LOLD events as more wind capacity is installed (Sheehy et al., 2016) despite LOLE/EEU remaining constant, but here we have showed these results are robust to a variety of choices regarding experimental configurations (see Table 5.2); the survival function of EU for the rest of the tested models can be seen in Appendix D.2.

5.6.2.4 A progression to a regime of fewer days of much more severe shortfalls

Another important change in the system's risk profile as more wind capacity is installed is the progression toward a system with fewer days of more severe shortfalls. Even more specifically, one where hourly loss of load events tend to occur in fewer but longer runs, defining a run as a sequence of one or more consecutive hourly time indices. To see this more clearly, we make use of the loss of load count (LOLC) statistic as defined in (Tindemans and Strbac, 2015),



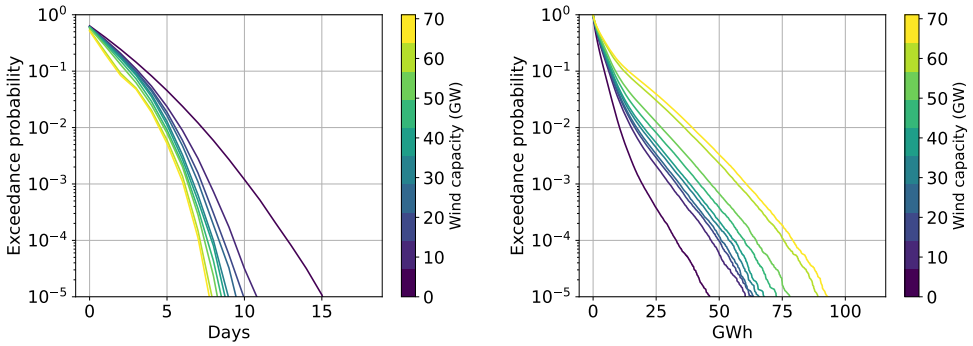
(a) LOLE = 3 hrs/yr configuration (b) EEU = 3 GWh/yr configuration

Figure 5.4: Survival function for EU (in log scale) in two experimental configurations with reliability standards of LOLE of 3 hrs/year and EEU of 3 GWh/year respectively.

which counts the number of times the capacity surplus process Z_t in (2.3) crosses from non-negatives to negatives values in the period of interest; in essence, this metric counts the number of loss of load runs as defined above. For the discrete hourly capacity surplus model we have used this can be computed as

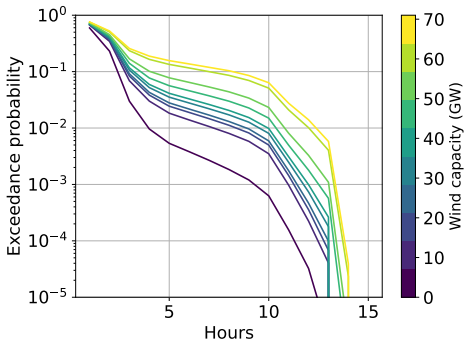
$$\text{LOLC} = \mathbb{I}(Z_1 < 0) + \sum_{t=2}^{n} \mathbb{I}(Z_{t-1} \geq 0) \cdot \mathbb{I}(Z_t < 0) \tag{5.12}$$

Figure 5.5 shows the survival functions for different risk metrics relevant to this change in system behaviour; we see a sharp and progressive decrease in LOLC for systems with wind penetration in Figure 5.5d compared to a baseline of no wind, which provides evidence that fewer loss of load runs occur at high wind regimes. However, as the LOLE is fixed across wind capacities, this also means loss of load runs are longer on average. The number of days with shortfalls also shows a sharp decrease, but these fewer days have considerably more severe shortfalls as measured by overall EU and LOLD, as shown by Figures 5.5b and 5.5c, where their relative frequency increases by multiple orders of magnitude for very high wind capacities compared to a baseline of no wind capacity. For instance, when conditioning to days with capacity shortfalls, events of LOLD $>$ 9hours occurred with a probability of roughly 10^{-3} for a system with no wind capacity, while for a system with 71 GW of installed capacity this increases to roughly 0.1, i.e., once every 10 days with shortfalls on average. In the case of energy unserved (again conditioned to days with capacity shortfalls), events with EU $>$ 16 GWh occur with a probability of roughly 10^{-3}, which becomes roughly 0.12 for a system with 71 GW of wind capacity, e.g. once every eight days with shortfalls on average. Note that for a system with such a high level of wind capacity, the frequency of

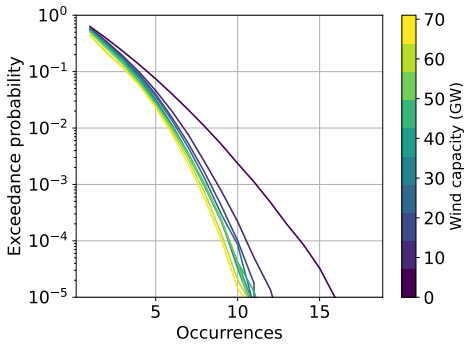


(a) Number of days with shortfalls

(b) EU in days with shortfalls



(c) LOLD in days with shortfalls



(d) LOLC

Figure 5.5: Survival functions for risk metrics related to the number of days with shortfalls and their severity, as well as LOLC. These results correspond to the standard experimental configuration (Table 5.2), but results were consistent across configurations.

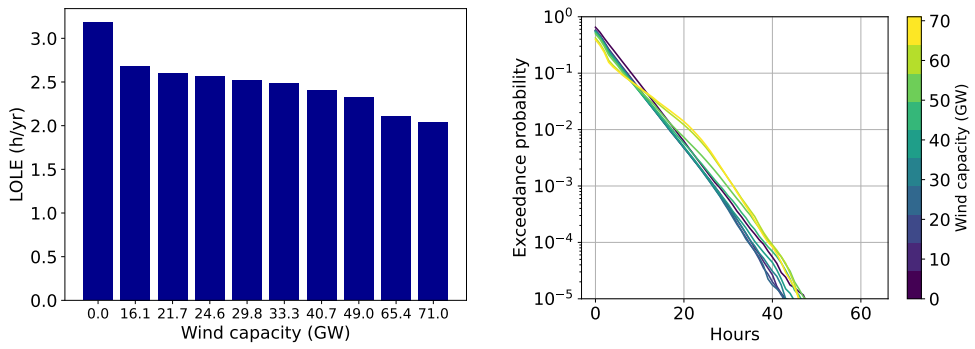
days with capacity shortfalls also decreases considerably (Figure 5.5a), however, this again suggests that when shortfalls do occur, they are potentially much more severe than in low wind regimes.

A similar progression in the system's risk profile as shown by Figure 5.5 was observed across all of the tested experimental configurations as more wind capacity was installed. Among other things, this means this progression occurs even without accounting for the high-risk year of 2005. A possible reason for this change is the fact that an increasing reliance on wind exposes the system to shortfalls induced by low-wind weather patterns, which could last for multiple hours or more. From the point of view of decision-makers, the progression to a system with fewer days of potentially much more severe shortfalls, and an overall pattern of fewer but longer runs of hourly shortfalls is likely to be an important consideration; nevertheless, these trends are again not reflected by long-term averages such as LOLE or EEU.

5.6.2.5 Some results for a system with EEU-based reliability standards

Lastly, we offer a brief discussion of some interesting results for the experimental configuration with a fixed EEU of 3 GWh/year; this particular normalising value was chosen based on the EEU for a system with no wind capacity and a LOLE of 3 hrs/year, which was 2.9 GWh/year. The LOLE progression and LOLD survival function for this configuration across wind capacities are shown in Figure 5.6. We can see that for this reliability standard, the system's LOLE decreases as more wind is installed (unlike our results for a system with a LOLE-based reliability standard where EEU consistently increases for progressively larger wind capacities).

Although the evolution of the LOLD distribution as more wind is installed is similar under both LOLE and EEU-based reliability standards, under the latter, the increase in the frequency of severe events (Figure 5.6b) is smaller than under the former (Figure 5.3a). However, the general points made above about the increase in frequency of severe EU events, as well as a regime of fewer days with more severe shortfalls and fewer runs of hourly loss of load events, were also valid for this experimental configuration (see Figure D.5). However, besides entailing significantly larger amounts of capacity to procure in order to meet such a reliability standard, it might also conflict with the underlying economic rationale of the capacity procurement methodology in GB, as the latter assumes there is an optimal fixed level of LOLE that minimises costs (Zachary et al., 2021), while Figure 5.6a shows the LOLE of a system with a fixed EEU consistently decreases as more wind is installed.



(a) LOLE (b) LOLD

Figure 5.6: Results for LOLE and LOLD in a system with the fixed EEU configuration (Table 5.2).

5.6.2.6 Summary

In this Section we have illustrated concrete changes in the system's risk profile that would be of interest for (risk-averse) decision-makers as more wind capacity is installed, namely, much higher year to year variability in outturn (Section 5.6.2.1) with more frequent occurrences of severe EU events (5.6.2.3), fewer days with more potentially much more severe shortfalls and a tendency of loss of load events to occur in fewer but longer hourly runs (Section 5.6.2.4), defining a run as a sequence of one or more consecutive hourly time indices. Overall, for a system with very high levels of wind capacity this suggests a picture of a lower number of years and days with capacity shortfalls but where shortfalls that can be potentially much more severe, which is no doubt an important consideration for policy makers. Moreover, we have also illustrated that a high reliance on renewable resources introduces substantial sampling uncertainty into the results and interpretation of adequacy calculations due to the limited number of historic years available to evaluate future possible weather conditions (Section 5.6.2.2). Lastly, we explored some differences in results for a system with a reliability standard based on EEU (5.6.2.5).

Finally, we note that events with probabilities less than, say, 10^{-3} in these experiments are likely unrepresentative of the real system, as there is a high level of parametric and epistemic uncertainty in these models. However, these experiments do illustrate the more general trends noted above as the systems transitions to a very high reliance on wind capacity. Moreover, these trends seem to be present for a variety of system configurations (see Table 5.2), which suggests these results may generalise beyond the data we have used here.

5.7 Sensitivity analysis of optimal procured capacity

We now analyse the problem of capacity procurement in GB from the point of view of risk-averse decision-makers. The objective of this Section is to incorporate risk aversion into the current decision analysis framework (Section 5.3), and to look at how misspefication of the degree of risk aversion and the VOLL parameter affect the output of this process in the context of an increasing reliance on wind. We do not explore the effects of these parameters on the wider LWR solution in (5.8), and limit the analysis to the individual cost functions of any given future scenario in the sense described in Section 5.3

5.7.1 Incorporating risk aversion into the cost function

As discussed previously in this Chapter, decision-makers dealing with the possibility of events with very severe consequences, as can be the case with capacity shortfalls, might exhibit a degree of risk aversion. Nevertheless, this is not presently accounted for by the current methodology in GB. We incorporate this by replacing the measure of the system's unreliability, namely the EEU(x) term in (5.7),by the conditional value at risk (CVaR) of the energy unserved. To illustrate this, we first define the value at risk (VaR) for a random variable X and a probability level $\alpha \in [0, 1]$ as

$$\text{VaR}(X; \alpha) = \min \{x \in \mathbb{R} \mid F_X(x) \geq \alpha\}$$

where F_X is the cumulative distribution function of X. Thus, $\text{VaR}(X; \alpha)$ is just the α-level quantile of X. We can now define CVaR as done in (Sarykalin et al., 2008) by writing

$$\text{CVaR}(X; \alpha) = \int_{-\infty}^{\infty} z dF_X^{\alpha}(z) \tag{5.13}$$

where

$$F_X^{\alpha}(x) = \begin{cases} 0, & \text{when } x < \text{VaR}_{\alpha}(X) \\ \frac{F_X(z)-\alpha}{1-\alpha}, & \text{when } x \geq \text{VaR}_{\alpha}(X) \end{cases}$$

Note that for continuous distributions (i.e., where there are no points with probability mass), (5.13) reduces to

$$\text{CVaR}(X; \alpha) = \mathbb{E}[X \mid X > \text{VaR}(X; \alpha)] \tag{5.14}$$

which is the expected value of X conditioned on an exceedance of its α-level quantile. However, this identity is not true for more general distributions, and this distinction is important here since the distribution of EU, defined in (5.9), has probability mass at zero, as peak seasons with no capacity shortfalls (i.e., EU = 0) have a positive probability of occurrence. For this reason, using (5.14) directly on the EU distribution would result in a discontinuous risk function with

respect to α. However, the more general definition of CVaR in (5.13) avoids this problem altogether; later in this Section, we show an explicit formula for the CVaR of the EU distribution. Lastly, we note that the α parameter in both VaR and CVaR can be seen as the degree of risk aversion from the decision-maker, with higher values representing an increasing degree of risk aversion. On the other hand, CVaR recovers the (risk-neutral) expectation $\mathbb{E}[X]$ for $\alpha = 0$.

The CVaR function has already been explored in optimisation problems related to system adequacy for incorporating a notion of risk aversion into the problem, for example, (Mays et al., 2019) analyses the implications of capacity market mechanisms on the decisions of risk-averse private investors. Another example is (Dias and da Costa, 2018), which formulates iterative algorithms for the problem of capacity expansion planning, using CVaR and the EU distribution in the reliability constraints, which are meant to account for the degree of risk aversion from stakeholders. We note that the latter referenced work is similar to this work in the sense that both are modelling societal investors, i.e., national governments that need to decide how much capacity to procure. On the other hand, (Mays et al., 2019) models private investors that participate in capacity markets. However, the principles of utility theory on which the use of CVaR is based are the same in both cases. Finally, CVaR is widely used in these and other fields of application partly due to the fact that it is a coherent risk measure, meaning it meets certain theoretical properties that are desirable in a risk function from the point of view of utility theory (Artzner et al., 1999). In particular, CVaR is a convex function in the space (X, α), where the first component denotes the space of random variables and α denotes the risk-aversion level; see (Sarykalin et al., 2008) for more details. This property makes CVaR particularly attractive in optimisation problems like this work and the applications mentioned above.

We now show the derivation of the CVaR function for the EU distribution that we use in the rest of this Chapter and the corresponding cost function for the capacity procurement problem. First, note that (5.6) can be expressed as

$$\mathrm{EEU}(x) = \mathbb{E}[\mathrm{EU}(x)] \tag{5.15}$$

$$\mathrm{EU}(x) = \sum_{t=1}^{n} (Y_t - X_t - x)_+ \cdot \Delta_t \tag{5.16}$$

We use (5.16) to directly replace $\mathrm{EEU}(x)$ in (5.7) by

$$\mathcal{R}(x; \alpha) = \mathrm{CVaR}(\mathrm{EU}(x); \alpha) \tag{5.17}$$

In order to derive a more explicit expression for (5.17), let us define

$$\begin{aligned} p_0(x) =& \mathbb{P}(\mathrm{EU}(x) = 0) \\ \mathrm{EU}^+(x) =& \mathrm{EU}(x) | \mathrm{EU}(x) > 0 \\ \mathrm{EEU}^+(x) =& \mathbb{E}[\mathrm{EU}^+(x)] \end{aligned}$$

Here, p_0 is the probability of having a peak season without capacity shortfalls, EU^+ is the distribution of EU conditional on being positive (i.e., conditional on peak seasons with capacity shortfalls), and EEU^+ is the expectation of the latter, and all of these quantities are parameterised by the level of procured capacity x. Then, we can write (5.17) as

$$\mathcal{R}(x;\alpha) = \begin{cases} \mathbb{E}\left[\mathrm{EU}(x) \mid \mathrm{EU}(x) > \mathrm{VaR}(\mathrm{EU}(x);\alpha)\right], & \alpha > p_0(x) \\ (1 - p_0(x))/(1-\alpha) \cdot \mathrm{EEU}^+(x), & \alpha \leq p_0(x) \end{cases} \tag{5.18}$$

This can be seen as the expected value of an EU outturn above a level which is exceeded every $(1-\alpha)^{-1}$ years on average. The corresponding cost function can be written as

$$c'(x;\alpha) = \mathrm{CONE} \cdot x + \mathrm{VOLL} \cdot \mathcal{R}(x;\alpha) \tag{5.19}$$

Note that (5.19) reduces to the original cost function (5.7) for $\alpha = 0$, this is, $\mathcal{R}(x;0) = \mathrm{EEU}(x)$, which is just a consequence of the fact that $\mathrm{CVaR}(X;0) = \mathbb{E}[X]$ for any random variable X.

Lastly, we note that it could be argued that the degree of risk aversion from the decision-maker would also affect the value of the VOLL parameter in the capacity procurement problem. The current methodology for estimating the VOLL relies on measuring the average costs across consumers per lost unit of energy, and in this sense it is a risk-neutral measure of the costs. Higher degrees of risk aversion could mean decision-makers place more weight on the costs of customers that are more severely affected by loss of load events, resulting in a larger VOLL. However, it is not clear how to properly incorporate this consideration into the formulation without additional data on consumer costs. For this reason, in this work we keep the effect of the risk aversion parameter limited to the CVaR function of the EU, and treat this experiment as a first approximation to this problem.

5.7.2 Experimental setup

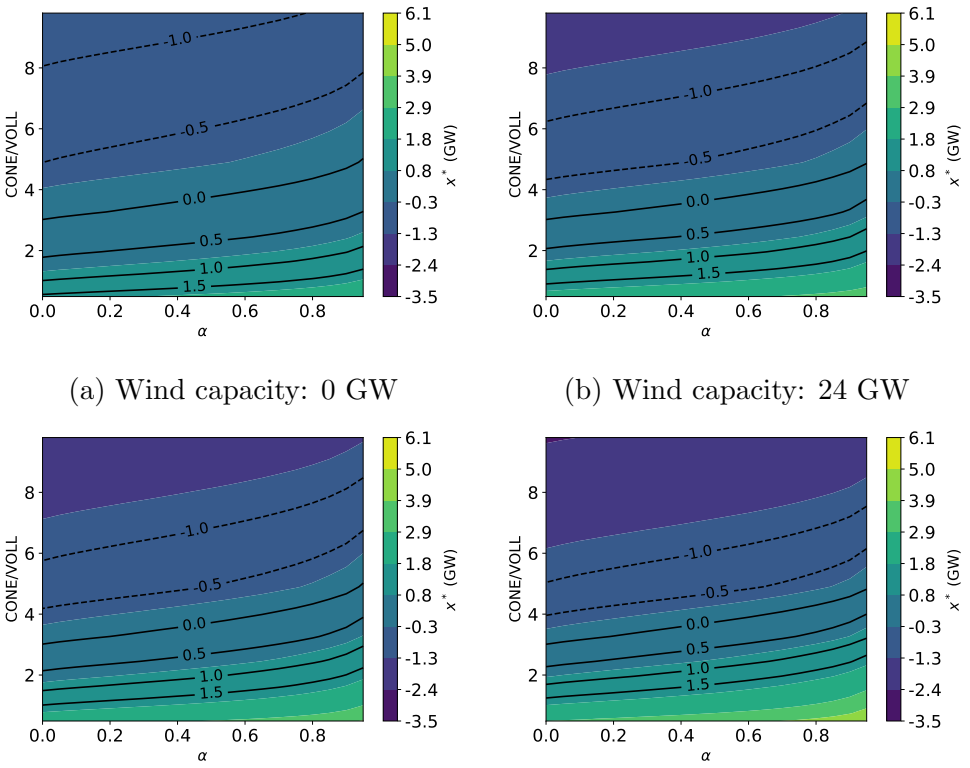
In the following Section, we present various visualisations that explore how the optimal procured capacity x^*, which minimises (5.19), depends on risk aversion, the VOLL parameter and the amount of installed wind capacity. We set a CONE of 49,000 £/MW-year, which is slightly above the CONE used by National Grid in (DECC, 2013c) to emphasise our model intends to merely be representative of the GB system. In general computing x^* for $\alpha > 0$ requires sequential simulation, as quantiles of the EU distribution cannot be calculated as outputs of a time - collapsed model. Moreover, numerical optimisation algorithms require multiple evaluations of the cost function (5.19), each of which require a pass through all of the simulated data, and so this is also more computationally expensive than the experiments in the previous Section which required a single pass through the data. For this reason results in this Section use a smaller simulated data set with 10,000 realisations of the same set of peak seasons used in the previous Section. To devise scenarios for installed wind capacity, we proceed analogously to the previous Section to produce 4 installed wind capacity scenarios (one of which is a baseline with no installed wind capacity), which are shown in Table D.2. For each one of these wind scenarios, (5.19) was minimised for a parameter grid (α, VOLL) of size 20×20; we use the system model corresponding to the standard experimental configuration in Table 5.2 to perform these experiments. All of this took approximately 14 hours using the same hardware configurations as in the previous Section.

Lastly, we note that different values of VOLL act on x^* through a ratio of the form CONE/VOLL in (5.19), in the sense that usual optimality conditions for (5.19) imply

$$-\frac{\partial \mathcal{R}}{\partial x}\Big|_{x=x^*} = \frac{\text{CONE}}{\text{VOLL}} \tag{5.20}$$

We hereafter call the right hand side the *cost ratio*, and note that it is just ratio between the procurement and unreliability costs per unit of energy under the current methodology in GB. In what follows we present and interpret the results in terms of this ratio rather than raw VOLL values, although the relationship between the two quantities is straightforward to see; moreover, analogous results in terms of raw VOLL can be found in Section D.3.

Lastly, in the following experiments we only compare different degrees of risk aversion through the values of optimal procured capacity or economic costs they produce in our formulation. However, in a more general sense, we note that direct comparison between different degrees of risk aversion, e.g. directly comparing



(a) Wind capacity: 0 GW

(b) Wind capacity: 24 GW

(c) Wind capacity: 41 GW

(d) Wind capacity: 71 GW

Figure 5.7: Contour lines for the optimal procured capacity x^* as a function of the risk aversion parameter α (X axis) and the ratio of CONE/VOLL (Y axis), for each wind capacity scenario. Dotted lines denote negative contour levels.

$\mathcal{R}(x;\alpha)$ for different values of α, might be conceptually problematic even though they are expressed in the same units, as in this case different values may be seen as different subjective assessments of the world by the decision-maker.

5.7.3 Results

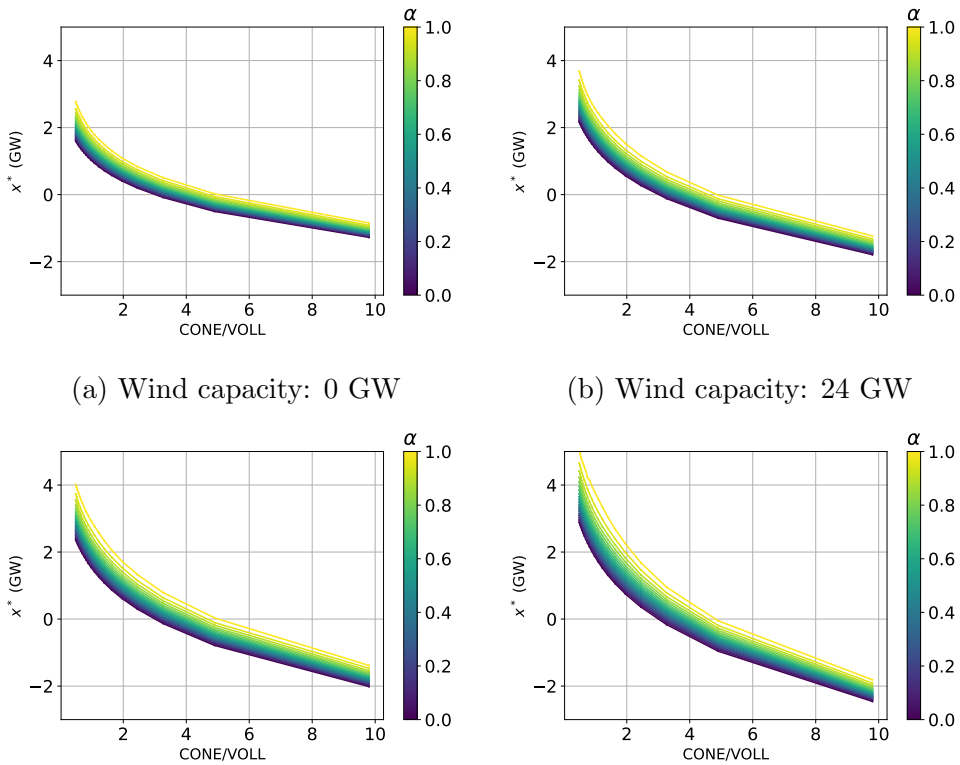
5.7.3.1 Sensitivity to cost ratio is higher than to risk aversion degree

Figure 5.7 shows results for procured capacity x^* as a function of both the cost ratio (Y axis) and α (X axis)[4]. For ease of comparison, the Figure's colormaps have been standardised to the same range and labelled contour lines show the same levels in all subfigures. For the systems from Figures 5.7a to 5.7d, the parameter $p_0 = \mathbb{P}(\text{EU} = 0)$ was 0.36, 0.38, 0.4 and 0.48, respectively; this means higher wind capacities produce less years with shortfall occurrences on average,

[4]Note that the normalisation to a LOLE of 3 hrs/year that we have performed fixes the contour line corresponding to $x^* = 0$ to a value close to 3 in the Y axis. The reason is that for $\alpha = 0$, it can be shown that (5.20) implies LOLE = CONE/VOLL at x^* (Zachary et al., 2021).

as was discussed in the previous Sections.

We can see that for all wind capacities the contour lines of x^* are almost linear, with a mild upward slope, except for very high values of α; an implication of this is that the procured capacity x^* is much less sensitive to changes in α than to changes in in the cost ratio, except for very high values of α. For example, in the case of no wind capacity, parameter values of $\alpha = 0.5$ and CONE/VOLL $= 3$ produce an x^* of approximately 200 MW, which is similar to the solution for $\alpha = 0$ and CONE/VOLL $= 2.6$, i.e., in terms of the procured capacity x^*, an increase in risk aversion from a risk-neutral value of $\alpha = 0$ to $\alpha = 0.5$ is equivalent to a decrease in the CONE/VOLL ratio from 3 to 2.6 , or an increase of around 2,000 £/MWh in VOLL (since CONE is fixed to £49,000 throughout these experiments).



(a) Wind capacity: 0 GW

(b) Wind capacity: 24 GW

(c) Wind capacity: 41 GW

(d) Wind capacity: 71 GW

Figure 5.8: Optimal procured capacity x^* (Y axis) as a function of the cost ratio (X axis) for each of the wind capacity scenarios.

5.7.3.2 Sensitivity to cost ratio grows as more wind is installed

The general shape of the contour lines do not seem to be altered across wind capacities in Figure 5.8, meaning sensitivity of x^* to α across wind capacities

remain roughly similar. However, sensitivity of x^* to the cost ratio increases substantially as wind capacity grows: looking at the scale of contour lines across the Y axis in Figure 5.7, we can see this sensitivity becomes larger as more wind is installed.

A clearer picture of this effect can be seen in Figure 5.8, which shows the procured capacity x^* (Y axis) as a function of the cost ratio (X axis) for the four scenarios of installed wind capacity. The Y axis range has been standardised to the same range for ease of comparison.

The increase in scale of procured capacity values x^* is evident from this Figure as the reliance on wind grows. Note that while the impact of the degree of risk aversion from the decision-maker α on x^* is also somewhat larger in high wind regimes, this effect is much smaller than that of the cost ratio. Sensitivity to α in Figure 5.8 seem to be largest when the cost ratio is small (i.e., large VOLL), but this is much less pronounced for large values of it (i.e., small VOLL).

Part of the reason for the increase in sensitivity to the cost ratio for higher wind capacities is likely the increase in tail heaviness of the EU distribution as the amount of installed wind capacity grows (Figure 5.4a). More specifically, as reliance on wind grows, heavier EU tails make CVaR values larger for any given $\alpha \in [0, 1]$, and this change in scale might in turn produce a change in scale in the values of optimal procured capacity through (5.20), ultimately exacerbating its sensitivity to the cost ratio.

5.7.3.3 Pareto frontiers of optimal costs are close to linear

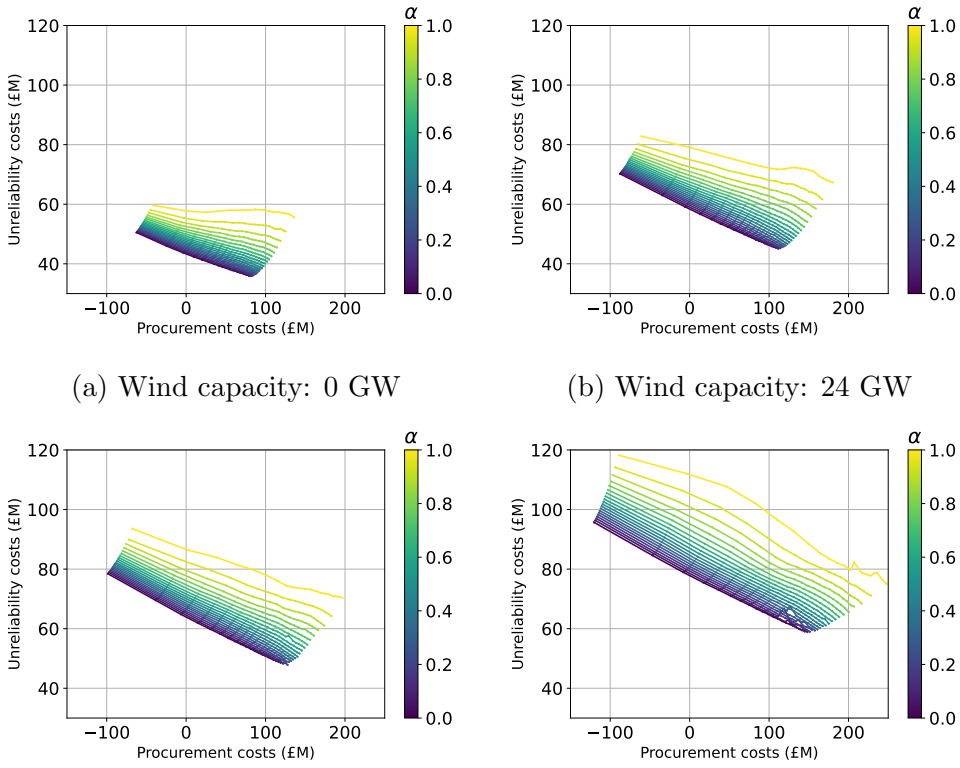
Lastly, in Figure 5.9, we show the Pareto frontier of procurement and unreliability costs for different choices of α and cost ratios. The displayed trajectories are the image of the functions $\kappa_\alpha : \mathbb{R}^+ \to \mathbb{R}^2$ where

$$\kappa_\alpha \left(\frac{\text{CONE}}{\text{VOLL}} \right) = \begin{pmatrix} \text{CONE} \cdot x^* \\ \text{VOLL} \cdot \mathcal{R}(x^*; \alpha) \end{pmatrix} \tag{5.21}$$

This is, κ_α maps the cost ratio to the optimal procurement and unreliability costs (note that since we keep CONE fixed, κ_α is a function of VOLL only). Larger values of the cost ratio (lower values of VOLL) correspond to lower procurement costs, i.e., the left end of the shown trajectories. It is clear from Figure 5.9 that higher wind capacities produce a positive change in location and scale of the optimal unreliability cost trajectories; this is, optimal unreliability costs become larger in general, but also more sensitive to the choice of cost ratio. On the other hand, a large positive change in scale is also observed for optimal procurement

cost trajectories, which become considerably more sensitive to the cost ratio as more wind is installed. Note that no change in location is observed in this case, e.g. optimal procurement cost trajectories remain centred at zero, which is caused by the normalisation to an LOLE of 3 hrs/year for all wind capacity scenarios. Finally, it is also clear that for higher wind capacities the optimal procurement costs are more sensitive to α when the cost ratio is low (i.e., right end of the trajectories); all these effects are in line with what was observed in previous Figures.

It is interesting to note that cost trajectories are fairly linear for virtually all parameter combinations (despite some visible numerical instability in Figure 5.9d), except for the case of high values of α in the baseline scenario of no wind capacity. In the latter case, optimal unreliability costs plateau around £60M for a VOLL of 15,000 £/MWh, and is virtually unaffected by larger VOLL values up to 100,000 £/MWh and their corresponding increases in procured capacity, i.e., in this case additional procured capacity prevents unreliability costs from growing, rather than decreasing them.



(a) Wind capacity: 0 GW

(b) Wind capacity: 24 GW

(c) Wind capacity: 41 GW

(d) Wind capacity: 71 GW

Figure 5.9: Bivariate trajectories of optimal procurement (X axis) and unreliability costs (Y axis) as a function of the cost ratio

5.7.3.4 Summary

Overall, in these experiments we observe that optimal procured capacity is significantly more sensitive to the cost ratio than to the degree of risk aversion of the decision-maker (Section 5.7.3.1). Moreover, as more wind is installed x^* becomes substantially more sensitive to the value of the cost ratio (i.e., CONE/VOLL), with sensitivity to α also increasing but to a lesser degree (Section 5.7.3.2). Finally, we show that the Pareto frontiers of the optimal costs for almost all parameter combinations are linear (Section 5.7.3.3).

A possible interpretation of these results is that (at least under the formulation we have used here) a principled specification of VOLL becomes more important as more wind is installed in order to procure appropriate a level of capacity that reflects the decision-maker's concerns. Here, the VOLL parameter should also reflect these concerns, since the decision-maker's degree of risk aversion would likely influence this parameter too, as per the discussion at the end of Section 5.7.1. In this sense, a principled specification of the decision-maker's degree of risk aversion would be required first in order to produce an appropriate VOLL parameter, even though in these experiments x^* is not very sensitive to α per se. However, we note that it remains to be explored to which extent the changes observed in these experiments translate to the output of the decision process, namely the LWR solution in (5.8).

We also note that, while these experiments are useful to illustrate the points outlined in the previous paragraph, the specific relationship between x^* and $\mathcal{R}(x^*;\alpha)$ would likely be different in a real system for some of the explored levels of procured capacity: as was mentioned in Section 5.3, for small values of additional capacity (compared to what is already installed in the system), shifting the capacity surplus distribution as we do in (5.6) and (5.16) provides a reasonable approximation, but this may not hold when additional capacity values are the order of multiple GWs. In the latter case, available additional capacity may be more accurately represented by a probability distribution, and so higher order moments of the total ACC distribution may also change when this is added to the system. In particular, it is likely our experiments overlook additional variance in ACC when adding multiple GWs of additional conventional capacity, which might mean values of $\mathcal{R}(x^*;\alpha)$ would be higher than observed in these experiments. This in turn could make the sensitivity of x^* to be even higher than what we observed in Figure 5.8.

5.8 Discussion

Overall, results in this Chapter are along the same lines as the discussion in (Stenclik et al., 2021): as power systems progress to a regime of very high renewable generation, adequacy risks are fundamentally altered, as most of the risk now comes from events of low renewable generation (which are dominated by weather patterns) rather than very high demand levels. Most of the changes that this transformation entails are not reflected by traditional risk indices like LOLE and EEU, even though these changes are important for system planning. Thus, a more comprehensive analysis of the system's risk profile that accounts for shortfall size, frequency and duration is needed in order to give decision makers the right tools to make principled choices.

In addition to this, robust methodology for accounting for uncertainty in future weather variability that is not reflected in historic data will be critical in adequacy calculations for power systems dominated by renewable generation. Approaches that complement analysis of historical data with appropriate probabilistic models for weather conditions outside of the historic record, as well as new modes of failure, are needed to properly address this problem.

Finally, we have shown that devising appropriate formulations for the capacity procurement problem that better reflect the concerns from decision-makers is going to be increasingly important as reliance on wind capacity grows in order to procure amounts of capacity in line with decision-maker's concerns, which may not be those of a risk-neutral agent.

5.8.1 Future work

The system model that we have used here does not take into account interconnection to other systems; interconnection will play an important role in offsetting the increased variability entailed by a higher reliance of intermittent sources like wind and solar (M. Grams et al., 2017), and so performing an analogous multi-area study that looks at the degree to which interconnection changes the picture described in this Chapter could be a direct extension to this work. For the two-area case, the software used in this Chapter (see Section A) can be readily used to that end. As more wind is installed in the coming decade (which is the time horizon we have looked at here), other technologies on both the supply and demand sides, such as energy storage and electric vehicles, might start to experience more widespread adoption. Exploring the effect of these technologies on the results we obtained would also be an important extension to this work.

Another possible extension is incorporating more detailed unreliability cost functions in the capacity procurement formulation that better reflect the concerns of decision-makers. For example, the degree of risk aversion from the decision maker is likely to also affect the VOLL, as more risk averse agents may also give more weight to customers that experience higher costs per value of lost unit of energy, resulting in larger VOLL parameters. Moreover, one might consider more complex unreliability cost functions than VOLL times a measure of energy unserved, for instance, one in which VOLL is an increasing function of the amount of energy unserved, reflecting the notion that costs of capacity shortfalls grow superlinearly with shortfall size. In this case, the VOLL itself would be calculated as a model output according to a specified function of the EU.

Results on the substantial increase in sampling uncertainty in the output of adequacy calculations for very high wind capacities suggest the need to develop appropriate ways to account for this. Accounting for events outside the historic record, which include completely new modes of failure, is a difficult problems as by definition there is no direct data available for analysis, and as such requires the use of expert judgement and elicitation. In this regard, principled probabilistic weather models elicited using insights from domain experts is one possibility. For instance, using graphical models, such as Bayesian networks, as a foundation for this, linking scientific evidence of long-term climate trends to their possible effects on renewable generation. Results in this Chapter suggest these and related avenues of research may be critical for the future of system adequacy modelling in GB.

Chapter 6

Conclusion

In this work we have developed statistical methodology for system adequacy modelling on top of some of the practices currently in use in Great Britain, and have use data from Great Britain (GB) and Ireland (IRL) to analyse risk for both single-area and two-area systems, usually in the context of an increasing reliance on interconnection and wind capacity.

Results from Chapter 3 suggest the reliability contributions of interconnection is highly dependent on both the shortfall-sharing policy and the risk metric used to measure this, which emphasises the importance of setting clear arrangements between system operators and carefully choosing an appropriate measure of risk for the problem at hand. We also showed that a share policy as described in (Tindemans et al., 2019) can be undesirable from the point of view of small systems, as this policy can make interconnection less valuable than under a veto policy, or even increase risk depending on interconnection size and the other system's background.

In Chapter 4, we built on the work from the previous chapter and developed a parametric bivariate model based on results from extreme value theory (EVT) to model the tails of net demand in a two-area system. This model showed high goodness of fit and produced visible differences in post-interconnection LOLE estimates when compared to hindcast models. We also showed that the latter produce risk estimates driven by a very small number of historic observations, which suggests sampling variability of its estimates may be considerable. Moreover, we used the concepts of asymptotic dependence/independence from EVT to characterise statistical association in net demands and capacity deficits across areas, finding capacity deficits have asymptotically independent tails, a much weaker type of dependence than in net demands. This difference is driven by the assumption of independence between conventional generation fleets across areas,

which is commonly made in adequacy studies. One practical implication of this is that the occurrence of simultaneous capacity shortfalls at both areas is very rare, despite simultaneous occurrence of extreme net demand levels being much more common.

Finally, in Chapter 5 we found multiple important changes in the risk profile of a single area system representative of GB as it progresses to very high levels of wind capacity. Nevertheless, these changes are not reflected by risk indices based on long-term averages such as LOLE and EEU, and only become apparent upon closer inspection of the probability distribution of different aspects of loss of load events. Namely, we have found a progression towards a system with extreme variability in year-to-year outturn which can be orders of magnitude larger or smaller than long-term average statistics, a lower number of years with capacity shortfalls at the expense of a considerable increase in the frequency of severe loss of load events, and fewer days of potentially much more severe shortfalls both in terms of LOLD and EU. Lastly, loss of load events tend to occur in fewer, longer runs, defining a run as a sequence of one or more consecutive hourly time indices. Another important change is the fact that a high reliance on wind introduces a substantial amount of uncertainty into the results of adequacy calculations, due to the fact that the limited number of historic years to perform these calculations cannot represent the full range of weather conditions that the system might experience.

Moreover, as more wind capacity is installed, the solution to the capacity procurement problem becomes much more sensitive to parameters such as the value of lost load (VOLL). Thus, as this progression takes place, devising principled estimate of VOLL, and in general, formulations of the capacity procurement problem that better reflect the concerns of decision-makers, will become increasingly important to procure amounts of capacity in line with the decision-maker's concerns, which might not be those of a risk-neutral agent.

6.1 Future work

EVT models and Monte Carlo estimation provide a mathematically convenient, and computationally scalable alternative to perform adequacy assessment in multi-area systems where the number of individual systems is large. Performing an analysis similar to Chapter 4 in a system with more than two areas could be a direct extension to this work. A graphical model structure like the one proposed in (Engelke and Hitz, 2018) could help reduce the complexity from high-dimensional

extreme value analysis, and semi-parametric EVT models proposed in (Heffernan and Tawn, 2004; Keef et al., 2013; Liu and Tawn, 2014) together with more classical parametric models like the one we have used in this work could also help characterising the wide range of tail dependence structure that may be found between pairs of individual systems.

A direct extension to the work from Chapter 5 would be the inclusion of interconnection into the model. Interconnectors will play a big role offsetting additional variability from wind and other intermittent resources (M. Grams et al., 2017), and it would be worth exploring how interconnection changes the picture we described in said Chapter. The inclusion in our model of new technologies that may be more widely adopted in the near term, such as grid-level battery storage and electric vehicles, would also be an important extension to this work. Lastly, we note that in Chapter 5, we did not look at how the wider least worst regret (LWR) solution to the problem of capacity procurement was affected by the model parameters on which we performed sensitivity analysis, and limited this analysis to the cost function of individual future scenarios. Investigating the effects of these parameters on the LWR solution is another possible extension to this work.

Another interesting direction may be investigating more general approaches to incorporating the degree of risk aversion of the decision-maker into the capacity procurement problem, or more generally, making the capacity procurement problem reflect the concerns of decision-makers as closely as possible. For instance, in Section 5.8.1 we have discussed the possibility of considering more complex function of energy unserved than in the current formulation. In this regard, one may consider the case where the cost per unit of energy unserved is an increasing function of the amount of energy unserved, reflecting the notion that the economic impact of capacity shortfalls grows super-linearly with shortfall size. In this case, VOLL may be specified as a function of energy unserved, and its outturn calculated as a part of the model's output. These and other alternative formulations for VOLL should ideally make it a function of the decision-maker's subjective considerations, e.g. its degree of risk aversion.

Finally, we saw that meteorological conditions outside of the historic record could produce very different results than those from adequacy studies performed on available historic data, for systems highly reliant on wind capacity (see Section 5.6.2.6). It is critical to address this issue in order to help decision-makers make principled choices. Elicitation and expert judgement are essentially the only tools to address this problem, and one possibility is the development of principled probabilistic climate models based on insights from domain experts, that

link evidence on long-term climate trends to their possible impact on renewable generation.

Appendix A

Dissemination of methods: `riskmodel` Python package

Throughout the course of this research project, a large amount of code was developed to perform various common tasks in adequacy procurement and run different kinds of experiments; this code was subsequently improved, refactored, tested and integrated into a coherent Python package that is publicly available through Python's PYPI repository and can be installed in the usual way using `pip`. Making well-tested, reproducible code that could be reused by other researchers was one of the main objectives throughout this PhD project.

This package encompasses the three main bodies of research outlined in this thesis, namely:

1. Exact two-area EEU and LOLE calculations under a time-collapsed hindcast models and for both a veto and share policies (see Chapter 3).

2. Univariate and bivariate functionality for extreme value models, including graphical diagnostics, logistic and asymmetric logistic dependence models and hypothesis testing for asymptotic dependence (see Chapter 4)

3. Univariate sequential models as used in Chapter 5. Although said Chapter only outlined results for a single-area model, analogous two-area models are also implemented with veto and share policies.

The code is also openly available on Github[1], and has an extensive API documentation that follows standard format guidelines[2]. In particular, point 2 above contributes to the implementation in Python of extreme value methodology, which at the moment is not widespread outside the R programming language.

[1] https://github.com/nestorSag/riskmodels
[2] https://nestorsag.github.io/riskmodels/

For more detailed information readers are encouraged to look at the API documentation or the Github repository. Below, a high-level overview of the package is given.

A.1 Computational efficiency

Particular attention was given to computational efficiency, and a number of C-based extensions included with the package were developed, in particular for points 1 and 3 above. For sequential model simulations, an efficient C implementation of Markov chain based available conventional generation (ACG) is available, which showed a boost in speed of about two orders of magnitude when compared with pure R or Python implementations. This is not limited to two-state chains as is usually done in adequacy studies, and in principle general Markov chain models that incorporate de-rated states or start-up states could also be used. During sequential simulations, the package scales computations through a (local) map-reduce processing pattern where multiple files of simulated ACG values are repeatedly processed in parallel; this allowed for computations using up to a million simulated years of ACG data to be performed in a reasonable time, depending on the number of passes needed; single passes over said data sets were done in a matter of seconds using the hardware as described in 5.6. However, we found that for scales larger than this (for instance, for performing computations from Section 5.7 using larger ACG data sets) a distributed approach is likely to be needed, as the local multi-core approach taken by the package at this time is not performant enough, due to storage and memory bandwidths becoming serious bottlenecks. This can in principle be fixed relatively easily by developing small extensions to make the package take advantage of mature, massively parallel processing frameworks such as Apache Spark.

A.2 Algebraic manipulation of discrete distributions

Discrete distributions are extensively used throughout adequacy studies, particularly for time-collapsed models, where both ACG and net demand are usually modelled as discrete variables. The package implements functionality to handle them in an algebraic manner when it comes to affine transformations, e.g. `z = a*x + b*y + c` would produce the correct distribution for `z` by convolving the appropriate probability mass functions (hence, this assumes they are inde-

pendent); exceedance functions are implemented through logical operators, e.g. `z > k` would produce the appropriate distribution for $Z|Z > k$; all of these features also work when one of the objects is a semi-parametric model with a generalised Pareto tail model, as is usually done in extreme value theory, since the package also handles the generalised Pareto mixture that would result in such a case; finally, general a `map` method is implemented for general transformations.

A.3 Implemented extreme value models

At the time of this writing, available extreme value models include:

1. Univariate MLE-based generalised Pareto tail models
2. Univariate Bayesian generalised Pareto tail models
3. Logistic bivariate MLE-based tail dependence
4. Asymmetric logistic bivariate MLE-based tail dependence

The package can work with univariate or bivariate empirical distributions based on observed data, and fit univariate or bivariate models easily. In the case of bivariate models, the marginal tail models and the dependence model are fitted separately, and at the moment joint optimisation is not supported.

Available diagnostics before and after model-fitting include:

1. Mean residual life plots
2. Goodness of fit plots for univariate and bivariate models including profile log-likelihood, likelihood surface, return levels and Q-Q plots
3. Empirical Pickands dependence function

The hypothesis test for asymptotic dependence used in Section 4.5.1.2 is also available.

Appendix B

Chapter 3 appendix

B.1 LOLE and EEU formulas for 2-area time-collapsed hindcast calculations

In this section, the closed form of LOLE and EEU used in numerical experiments of Chapter 3 are shown; we include some comments on the computational efficiency of these calculations

B.1.1 Computational considerations

Due to the shape of the relevant regions that have to be integrated in Figure 3.3, these formulas usually involve adding up probability mass on triangular sections of the integer lattice that form the support of the available conventional generation distribution; a naive approach that scans this section would have a quadratic complexity on the triangle's side lengths. By using arrays of CDF values instead of adding up probability mass from scratch, this can be reduced to linear complexity on the triangle's side length. However, we can do better than this by taking advantage of the fact that both of the triangle's sides are of the same lengths; this allows recursion to be used in the calculation, which achieves logarithmic computational complexity. This is done as follows.

Let $(x, y), (x+l, y), (x, y+l)$ be the vertices of a right triangle inscribed in an integer lattice with a side length $l > 0$. This region has a probability mass of

$$p(x, y, l) = \sum_{x_1=x}^{x+l} \sum_{x_2=y}^{y+l-x_1} f_{X_1}(x_1) f_{X_2}(x_2) \tag{B.1}$$

where f denotes the corresponding probability mass function. First note that the probability mass of the square inscribed in this triangle with side lenght $\lfloor l/2 \rfloor$ and

lower left corner (x, y) can be calculated in constant time using CDF evaluations

$$\begin{aligned} s(x, y, \lfloor l/2 \rfloor) =& F_{X_1}\left(x + \lfloor l/2 \rfloor\right) F_{X_2}\left(y + \lfloor l/2 \rfloor\right) + \\ & F_{X_1}\left(x\right) F_{X_2}\left(y\right) - \\ & F_{X_1}\left(x\right) F_{X_2}\left(y + \lfloor l/2 \rfloor\right) - \\ & F_{X_1}(x + \lfloor l/2 \rfloor) F_{X_2}\left(x\right) \end{aligned} \tag{B.2}$$

Then, we have the following recursive relationship

$$\begin{aligned} p(x, y, l) =& s(x, y, \lfloor l/2 \rfloor) + \\ & p\left(x, y + \lfloor l/2 \rfloor, l - \lfloor l/2 \rfloor\right) + \\ & p\left(x + \lfloor l/2 \rfloor, y, l - \lfloor l/2 \rfloor\right) \end{aligned} \tag{B.3}$$

Using (B.3) we can reduce computational complexity to $\mathcal{O}(\log c)$ when calculating LOLE for a given interconnector capacity c. Note that this does not work for EEU, as in that case of an expectation each element in the triangular lattice has to be weighted, and this necessarily requires a full scan which inccurs in quadratic computational complexity.

B.1.2 Formulas

In the following equations, we assume that $d_i - w_i > 0$ for area $i, i = 1, 2$. For ease of syntax let us define a truncated expectation for a random variable X as

$$\mathcal{E}_X(g(X), a, b) = \sum_{x=a}^{b} g(x) \cdot \mathbb{P}(X = x) \tag{B.4}$$

Note that $\mathcal{E}_X(X, -\infty, \infty) = \mathbb{E}[X]$. For any given policy, under a hindcast net demand model for a historic dataset with n points we have

$$\text{LOLE} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{P}(\boldsymbol{M}_1' < 0 | \boldsymbol{D} = \boldsymbol{d}_i, \boldsymbol{W} = \boldsymbol{w}_i; c) \tag{B.5}$$

$$\text{EEU} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}_{\boldsymbol{X}}[\max(0, -\boldsymbol{M}_1') | \boldsymbol{D} = \boldsymbol{d}_i, \boldsymbol{W} = \boldsymbol{w}_i; c] \tag{B.6}$$

where c is a fixed interconnection capacity and $\boldsymbol{M'}$ is the vector of post-interconnection power margins. Finally, let

$$v_i = d_1 - w_i,\ i = 1, 2 \tag{B.7}$$

be the net demand in each system. Below, the form of the summands in (B.5) and (B.6) are specified for the veto and share policies and for area 1 in terms of CDF evaluations and functions B.3 and B.4; calculations for area 2 are performed analogously with swapped indices.

B.1.3 Veto policy

$$\begin{aligned}\mathbb{P}(\boldsymbol{M}_1' < 0 | \boldsymbol{D} = \boldsymbol{d}_i, \boldsymbol{W} = \boldsymbol{w}_i; c) =& F_{X_2}(v_2)F_{X_1}(v_1) + \\ & F_{X_1}(v_1 - c) - \\ & F_{X_2}(v_2)F_{X_1}(v_1 - c) + \\ & p(v_2, v_1 - c, c)\end{aligned} \tag{B.8}$$

$$\begin{aligned}\mathbb{E}_{\boldsymbol{X}}[\max(0, -\boldsymbol{M}_1') | \boldsymbol{D} = \boldsymbol{d}, \boldsymbol{W} = \boldsymbol{w}; c, \mathcal{V}] = & F_{X_2}(v_2)\mathcal{E}_{X_1}(v_1 - X_1, 0, v_1) + \\ & \mathcal{E}_{X_2}(\mathcal{E}_{X_1}(v_1 + v_2 - X_1 - X_2, 0, v_1 + v_2 - X_2), v_2, v_2 + c) + \\ & (1 - F_{X_2}(v_2 + c))\, \mathcal{E}_{X_1}(v_1 - c - X_1, 0, v_1 - c)\end{aligned} \tag{B.9}$$

B.1.4 Share policy

$$\begin{aligned}\mathbb{P}(\boldsymbol{M}_1' < 0 | \boldsymbol{D} = \boldsymbol{d}, \boldsymbol{W} = \boldsymbol{w}; c, \mathcal{S}) =& F_{X_2}(v_2 - c)F_{X_1}(v_1 + c) + \\ & F_{X_1}(v_1 - c) - \\ & F_{X_2}(v_2 - c)F_{X_1}(v_1 - c) + \\ & p(v_2 - c, v_1 - c, 2c)\end{aligned} \tag{B.10}$$

To calculate $\mathbb{E}_{\boldsymbol{X}}[\max(0, -\boldsymbol{M}_1') | \boldsymbol{D} = \boldsymbol{d}, \boldsymbol{W} = \boldsymbol{w}]$ let us first define

$$\alpha(y) = \frac{d_1}{d_2}(y - d_2 + w_2 - c) - d_i + w_i - c \tag{B.11}$$

$$\beta(y) = \alpha(y) + 2\left(1 + \frac{d_1}{d_2}\right)c \tag{B.12}$$

These are straight parallel lines that delimit the region of non-saturated interconnector flow in conventional generation space (see Fig. 3.3a); by using the

regions in the referenced Figure to compute the EEU, the following equality can be derived.

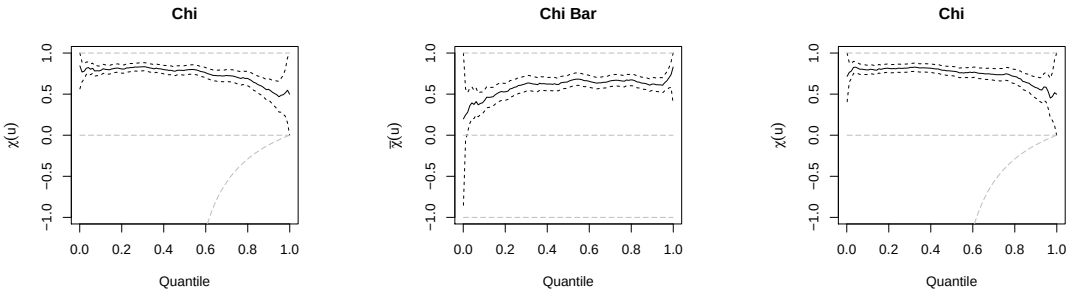
$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{X}}[\max(0,-\boldsymbol{M}_1')|\boldsymbol{D}=\boldsymbol{d},\boldsymbol{W}=\boldsymbol{w};c]= \\
&\mathcal{E}_{X_2}\left(\mathcal{E}_{X_1}\left(-\frac{d_1}{d_1+d_2}(X_1+X_2-v_1-v_2),\lfloor\alpha(X_2)\rfloor,v_1+v_2-X_2\right),v_2-c,v_2+c\right)+ \\
&\mathcal{E}_{X_2}\left(\mathcal{E}_{X_1}\left(-\frac{d_1}{d_1+d_2}(X_1+X_2-v_1-v_2),\lfloor\alpha(X_2)\rfloor,\lfloor\beta(X_2)-1\rfloor\right),0,v_2-c\right)+ \\
&\mathcal{E}_{X_2}\left(\mathcal{E}_{X_1}\left(v_i-c-X_1,0,\lfloor\alpha(X_2)-1\rfloor\right),\lceil v_2+c-d_2/d_1(v_1-c)\rceil,v_2+c\right)+ \\
&(1-F_{X_2}(d_2+w_2+c))\mathcal{E}_{X_1}\left(v_1-c-X_1,0,v_1-c\right)+ \\
&\mathcal{E}_{X_2}\left(\mathcal{E}_{X_1}\left(v_1+c-X_1,\lfloor\beta(X_2)\rfloor,v_1+c\right),0,v_2-c\right)
\end{aligned} \tag{B.13}$$

Note that some of the complexity in the above Equation stems from from the discretisation of conventional generation space into a lattice as well as from the change between interconnector saturation and non-saturation regimes delimited by sloped lines (B.11) and (B.12). Note that these lines are not always well aligned with said lattice.

Appendix C

Chapter 4 appendix

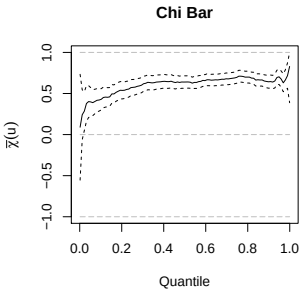
C.1 Additional figures



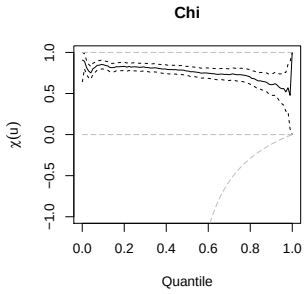
(a) 2007

(b) 2007

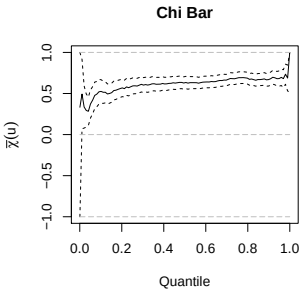
(c) 2008



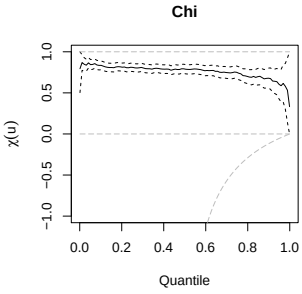
(d) 2008



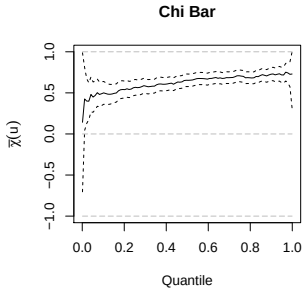
(e) 2009



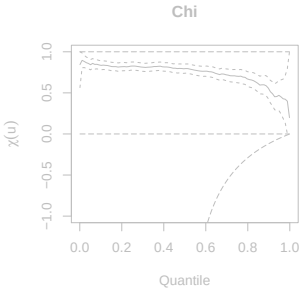
(f) 2009



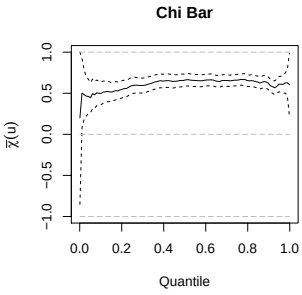
(g) 2010



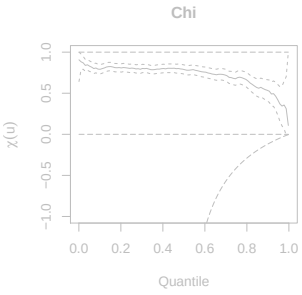
(h) 2010



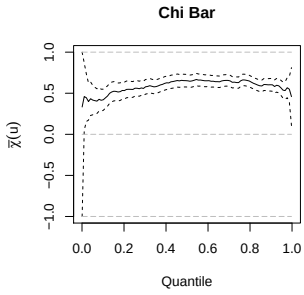
(i) 2011



(j) 2011



(k) 2012



(l) 2012

Figure C.1: Figures displaying estimates of χ and $\bar{\chi}$ for diagnosing the presence of asymptotic dependence (see C.2.3) in net demand data, as given by the `texmex` R package. Greyed out figures mean confidence intervals in the figure suggest there is no asymptotic dependence.

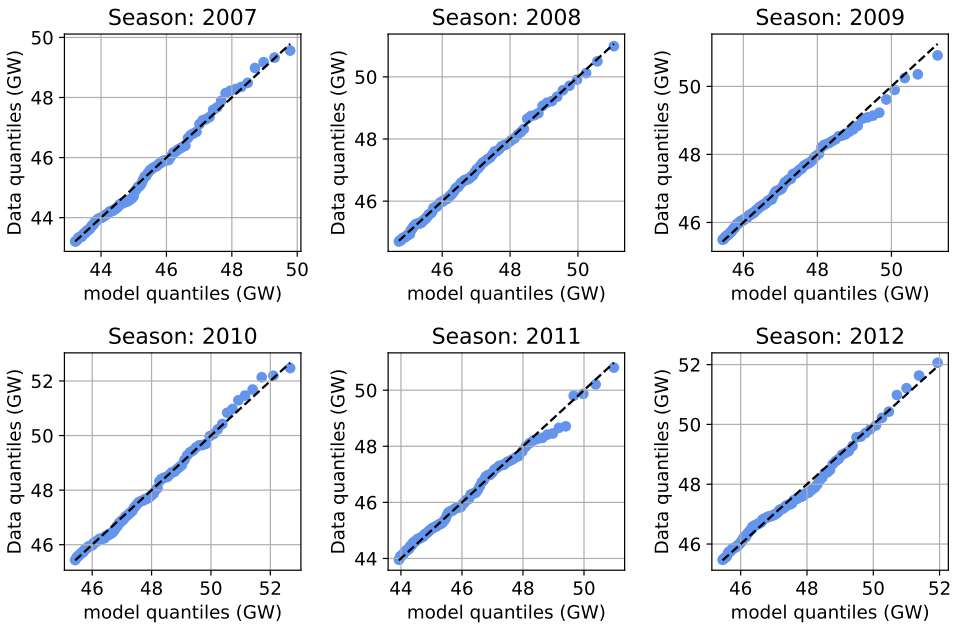


Figure C.2: Q-Q plots for fitted exceedance Generalised Pareto models on GB data

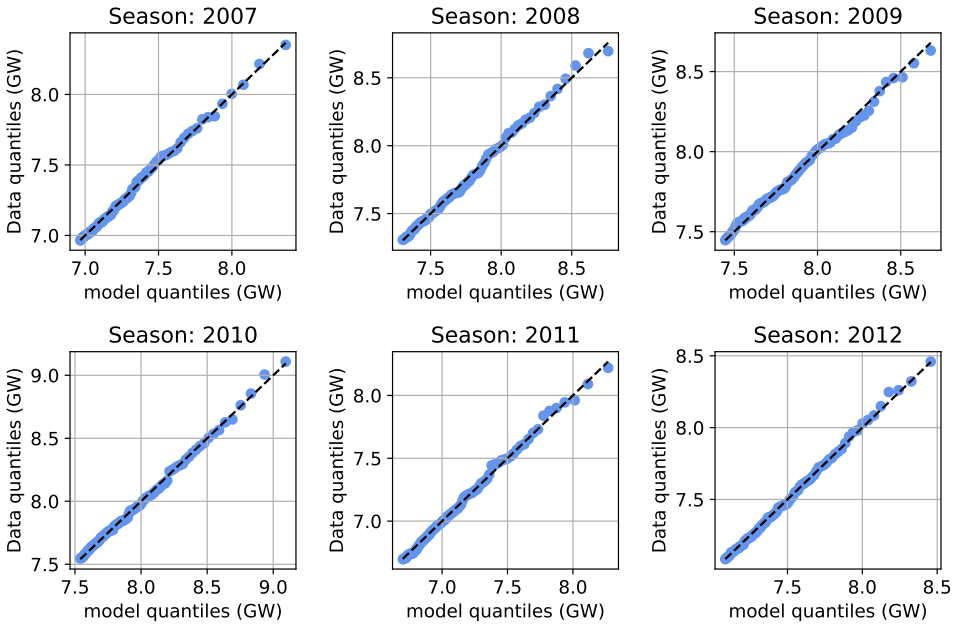


Figure C.3: Q-Q plots for fitted exceedance Generalised Pareto models on IRL data

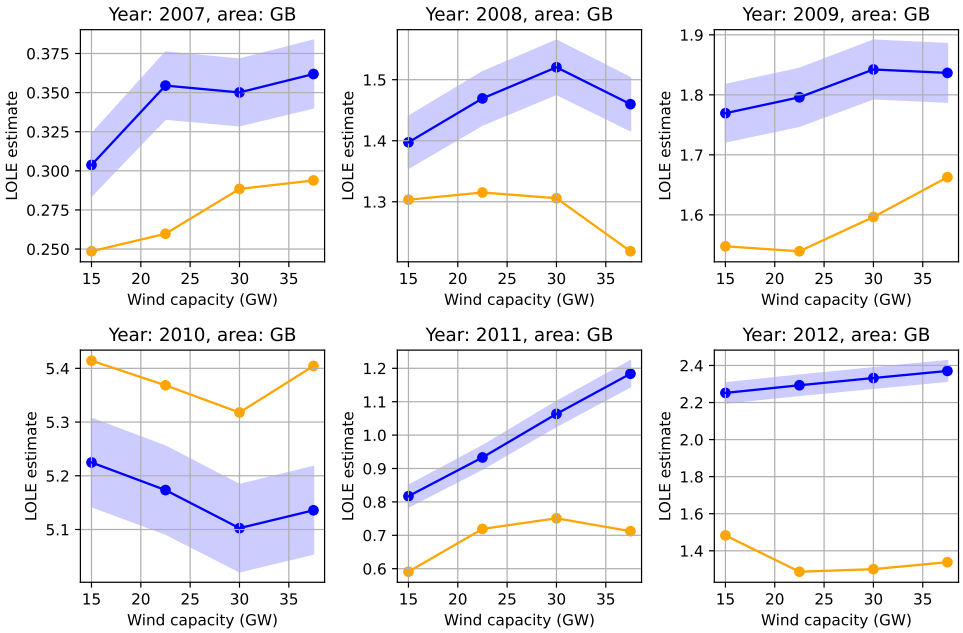


Figure C.4: Estimated post-interconnection values for LOLE in GB, following the experimental setup outlined in 4.5.2

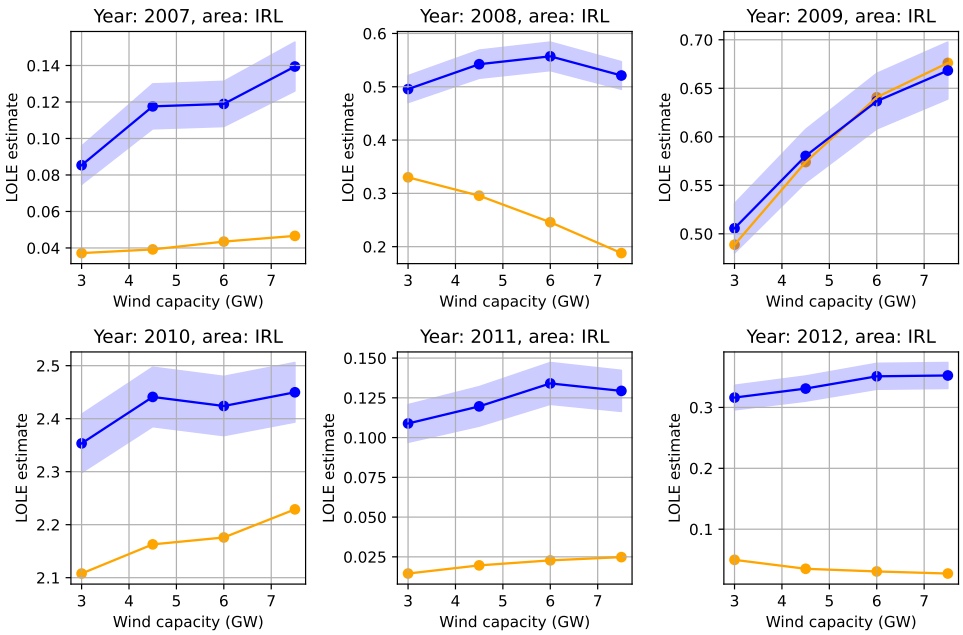


Figure C.5: Estimated post-interconnection values for LOLE in IRL, following the experimental setup outlined in 4.5.2

C.2 Review of extreme value theory

This Section offers a very brief introduction to the main results in univariate and multivariate extreme value theory that were used when working in Chapter 4. This is a condensed version of content that can be found in (Beirlant et al., 2006) and (Coles, 2013). Additional references are provided when needed.

C.2.1 Univariate EVT

Extreme value theory is concerned with estimating the distribution of sample maxima. Let us assume we have sample of independent, identically distributed (iid) values $x_1, , ..., x_n \in \mathbb{R}$ with probability density function (PDF) $f_x(\cdot)$ and cumulative distribution function (CDF) $F_x(\cdot)$. Let us also denote the sample maximum as $M_n = \max\{x_1, ..., x_n\}$. The extremal types theorem, which is arguably the most important result in univariate extreme value theory, characterizes the possible distributional limits of M_n

Theorem 1 (Extremal types theorem). *If there exist real sequences* $\{a_n\} \subset \mathbb{R}^+, \{b_n\} \subset \mathbb{R}$ *such that* $\lim_{n\to\infty} \mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq z\right) = G(z)$ *where* $G(\cdot)$ *is a non-degenerate distribution function, then* $G(\cdot)$ *belongs to one of the following families*

$$G(z) = \exp\left\{-\exp\left[-\left(\tfrac{z-\mu}{\sigma}\right)\right]\right\}, \quad -\infty < z < \infty$$

$$G(z) = \exp\left\{-\left(1 + \xi\left(\tfrac{z-\mu}{\sigma}\right)\right)^{\xi}\right\}, \quad \xi < 0, -\infty < z < -\xi^{-1}$$

$$G(z) = \exp\left\{-\left(1 + \xi\left(\tfrac{z-\mu}{\sigma}\right)\right)^{\xi}\right\}, \quad \xi > 0, -\xi^{-1} < z < \infty$$

where $\mu \in \mathbb{R}, \sigma \in \mathbb{R}^+$ *are location and scale parameters respectively.*

If the sample maxima of X converges in distribution to G, we say that F_X is in the domain of attraction of G. This is formalised in the following definition

Definition 1 (Domain of attraction). *Let* M_n *be the sample maxima of a distribution* F_X *such that*

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq z\right) = G(z) \tag{C.1}$$

for a non-degenerate distribution $G(z)$ and some sequences $\{a_n\} \subset \mathbb{R}^+, \{b_n\} \subset \mathbb{R}$. Then we say that F_X is in the domain of attraction of G, and denote it as $F_X \in \mathcal{D}(G)$.

The distributions in theorem 1 are called Gumbel, Weibull and Frechet distributions respectively, and for inference purposes the three families can be expressed as different subfamilies from a single distribution, called the generalized extreme value distribution (GEVD)

Definition 2 (generalized extreme value distribution)**.** *A random variable $X \in \mathbb{R}$ follows a generalized extreme value distribution with shape parameter $\xi \in \mathbb{R}$, scale parameter $\sigma \in \mathbb{R}^+$, and location parameter $\mu \in \mathbb{R}$ if its CDF is given by*

$$F_X(x) = \begin{cases} \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}, & \xi \neq 0 \\ \exp\left\{-\exp\left(\frac{x-\mu}{\sigma}\right)\right\}, & \xi = 0 \end{cases} \tag{C.2}$$

and we say that $F_X \sim GEVD(\mu, \sigma, \xi)$

The shape parameter ξ in C.2 controls the heavyness of the tails, with $\xi < 0$ giving rise to a distribution with a finite upper bound (Weibull distribution), $\xi = 0$ giving rise to asymptotically exponential tails (Gumbel distribution), and $\xi > 0$ to an asympotiticaly power-tailed distribution (Frechet distribution). These two results provide a justification for modelling extreme observation as following a GEVD, since if $(M_n - b_n)/a_n$ converges in distribution to $G(z) = \text{GEVD}(\mu, \sigma, \xi)$, as $n \to \infty$, then we can assume that for suitably large values of n

$$\mathbb{P}(M_n \leq z) = \mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq \frac{z - b_n}{a_n}\right) \approx G\left(\frac{z - b_n}{a_n}\right) = \text{GEVD}(a_n\mu + b_n, a_n\sigma, \xi) \tag{C.3}$$

Note that in (C.3), the mean and scale parameters of the GEVD approximation depend, in principle, on the sequences $\{a_n\}, \{b_n\}$; this is not an issue in practice.

Historically, estimation of the GEVD model was first done by arranging the data into multiple blocks, and using block maxima to perform maximum likelihood estimation. This is called the block maxima approach, and is more formally described in Algorithm 1.

Algorithm 1: Block maxima aproach GEVD estimation

Data: sample $x_1, ..., x_n$, number of blocks m

Result: estimated extreme value distribution GEVD$(\hat{\mu}, \hat{\sigma}, \hat{\xi})$

1 Partition the data into m blocks $b_1, ..., b_m$

2 set $y_i = \max\{x | x \in b_i\} \quad \forall i = 1, ..., m$

3 Estimate $\hat{\mu}, \hat{\sigma}, \hat{\xi}$ by maximum likelihood using $y_1, ..., y_m$ and (C.2)

In applications, it is common for the data to be indexed by time, and it is possible to group them by, say, year, in this way estimating the distribution of yearly maxima. If they are not indexed, it is also possible to shuffle them randomly into m different groups.

However, the block maxima approach is not very data-efficient. In this regard, a threshold exceedances approach, in which all data above a certain threshold is used for model fitting, is better, and it takes advantage of the following reasoning: assume, as for the block maxima approach, that the distribution of the maximum, M_n, in the data $x_1, ..., x_n$, is well approximated by a GEVD, i.e., $\mathbb{P}(M_n \leq u) \approx$ GEVD(μ, σ, ξ). Then we have

$$\mathbb{P}(M_n \leq u) = F_X^n(u) \approx \text{GEVD}(\mu, \sigma, \xi) = \exp\left(-\left(1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right)_+^{-1/\xi}\right)$$

then

$$n \log F_X(u) \approx -\left(1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right)_+^{-1/\xi}$$

If u is large enough so that $F_X(u) \approx 1$ we can use the Taylor expansion of logarithm around 1 (i.e., $\ln(x) \approx x - 1$ for $x \approx 1$) to get

$$\mathbb{P}(X > u) = 1 - F_X(u) \approx \frac{1}{n}\left(-\left(1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right)_+^{-1/\xi}\right)$$

which implies the following conditional probability distribution

$$\mathbb{P}(X > u + y | X > u) \approx \frac{\frac{1}{n}\left(-\left(1 + \xi\left(\frac{u+y-\mu}{\sigma}\right)\right)_+^{-1/\xi}\right)}{\frac{1}{n}\left(-\left(1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right)_+^{-1/\xi}\right)} = \left(1 + \xi \cdot \frac{y}{\sigma^*}\right)_+^{-1/\xi} \quad \text{(C.4)}$$

where $\sigma^* = \sigma + \xi(u - \mu)$; moreover, (C.4) becomes an equality in the limit as $u \to \infty$. We then say that the exceedances of X follows (asymptotically) a

generalized Pareto distribution, which is defined as below

Definition 3 (generalized Pareto distribution)**.** *A random variable $x \in \mathbb{R}$ follows a generalized Pareto distribution with shape parameter $\xi \in \mathbb{R}$, scale parameter $\sigma \in \mathbb{R}^+$, and location parameter $\mu \in \mathbb{R}$ if its CDF is given by*

$$F(x) = 1 - \left(1 + \xi \left(\frac{x-\mu}{\sigma}\right)\right)_+^{-1/\xi}, \; x \geq \mu \tag{C.5}$$

and we say that $F \sim GPD(\mu, \sigma, \xi)$

This reasoning suggest that instead of grouping the data into blocks, we can look at their exceedances over a high enough threshold u. This approach usually allows to use significantly more data, and Algorithm 2 the general procedure of estimation by threshold exceedances.

Algorithm 2: threshold exceedances approach GEVD estimation

Data: sample $x_1, ..., x_n$, threshold u such that $F_x(u) \approx 1$

Result: estimated extreme value distribution $\text{GPD}(u, \hat{\sigma}, \hat{\xi})$

1 create exceedances subset $S = \{x_i | x_i > u, i = 1, ..., n\}$

2 Use S to fit a generalized Pareto distribution by maximum likelihood

The threshold value u is a model parameter that has to be specified beforehand, and a typical way of doing so is looking at mean residual life plots, in the following way: note that if $Y \sim \text{GPD}(u, \sigma, \xi)$ with a shape parameter $\xi \leq 1$, then $\mathbb{E}[Y] = \sigma/(1-\xi)$. Now let us assume that for some sample of a random variable X, a generalized Pareto approximations hold well for thresholds of u_0 or more, then for $u > u_0$ we have

$$\mathbb{E}[X - u | X > u] \approx \frac{\sigma_u}{1-\xi} = \frac{\sigma_{u_0} + \xi u}{1-\xi} = \frac{\xi}{1-\xi} u + \frac{\sigma_{u_0}}{1-\xi}$$

where σ_u is the scale of the exceedances over u, so it follows that as we move the threshold upwards, the mean of the exceedances is a linear function of the threshold u. Hence, a mean residual life plot consists in plotting the function $h_u : (u, x_{\max}) \to \mathbb{R}$ such that

$$h_u(v) = \frac{1}{n_v} \sum_{i=1}^{n_v} \left(x_{(i)} - v\right)$$

where n_v is the number of samples above threshold v and $x_{(i)}$ are the corresponding sample indices; we should then select a threshold u over which $f_u(v)$ is approximately a straight line.

C.2.2 Multivariate EVT

In the multivariate case we study the component-wise maxima of multidimensional data. Let us assume we have a vector sample $\boldsymbol{x}_1, ..., \boldsymbol{x}_n \in \mathbb{R}^d$ with CDF $F_{\boldsymbol{X}}(\cdot)$ and define the vector relationship $\boldsymbol{x} \leq \boldsymbol{y}$ meaning $x_i \leq y_i$ for all $i = 1, ..d$, and the $\vee$ symbol as a maximum operator, e.g. $x \vee y = \max\{x, y\}$. We define the componentwise maximum of an n-sized sample, $\boldsymbol{M}_n$, as

$$\boldsymbol{M}_n = \left(\bigvee_{i=1}^{n} \boldsymbol{x}_{i,1}, ..., \bigvee_{i=1}^{n} \boldsymbol{x}_{i,d} \right)$$

Note that the component-wise maximum is not part of the data if maxima at different components were observed at different times. Although this fact may make $\boldsymbol{M}_n$ seem somewhat artificial, its analysis can give us valuable insights about the co-occurrences of extremes across components.

The CDF of $\boldsymbol{M}_n$ is $F^n_{\boldsymbol{X}}(\cdot)$, and we are interested in the limiting distribution of normalised component-wise sample maxima, such that for some suitable sequences $\{\boldsymbol{a}_n\}, \{\boldsymbol{b}_n\}$ we have

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\boldsymbol{M}_n - \boldsymbol{b}_n}{\boldsymbol{a}_n} \leq \boldsymbol{z} \right) = G(\boldsymbol{z})$$

where vector division applies component-wise. The univariate extremal types theorem (Theorem 1) ensures each of the marginal distributions of $G(z)$ is a GEVD with parameters $(\mu_j, \sigma_j, \xi_j), j = 1, ..., d$. We say $F_{\boldsymbol{X}}$ is in the domain of attraction of G, and denote this as $F_{\boldsymbol{X}} \in \mathcal{D}(G)$.

An important property of multivariate extreme value distributions $G(\boldsymbol{z})$ is max-stability, which is defined below

Definition 4 (max-stable distribution)**.** *Let $Z \in \mathbb{R}^d$ be a random vector with CDF $G(\boldsymbol{z})$, then G is said to be max-stable if for every $k > 0$, there exist $a > 0, b \in \mathbb{R}$ such that $G^k(\boldsymbol{z}) = G(a\boldsymbol{z} + b)$*

Intuitively, the sample maxima of max-stable distributions follows the same distribution with only a change of scale and location.

To avoid dealing with potentially many different distributions types at the margins, it is useful to map the components X_i so that they are identically distributed. A common choice is making them standard Frechet, so that their CDF is given by $G_j(z) = \exp(-1/z), j = 1, ..., d$; to do this, we can define

$$Z_j = -\frac{1}{\ln F_{X_j}(X_j)}, \quad 1 \leq j \leq d \tag{C.6}$$

Z_j follows a standard Frechet distribution, and since extreme value distributions are in their own domain of attraction, maxima from the transformed data will also follow a Frechet distribution, only with different location and scale parameters. The choice of margins is not important, and different choices can be made, for example, for Pareto distributions we can define

$$Z_j = -\frac{1}{1 - F_{X_j}(X_j)}, \quad 1 \leq j \leq d$$

and for Gumbel distributions we do

$$Z_j = -\ln(-\ln F_{X_j}(X_j)), \quad 1 \leq j \leq d \tag{C.7}$$

To simplify analytical calculations however, it is more convenient to deal with standard Frechet margins.

Let $G_*(\boldsymbol{z})$ be the limiting distribution of normalised maxima from the transformed vectors $\boldsymbol{z}_1, ..., \boldsymbol{z}_n$, then the standard Frechet margins and max-stability of G_* imply

$$G_*^k(k\boldsymbol{z}) = G_*(\boldsymbol{z}) \quad \forall k > 0, \boldsymbol{z} \in (\mathbb{R}^+)^d \tag{C.8}$$

moreover, it can be proved that every max-stable distribution G can be represented as

$$G(\boldsymbol{x}) = \exp(-\mu([\boldsymbol{q}, \boldsymbol{\infty}) \setminus [\boldsymbol{q}, \boldsymbol{x}])) \tag{C.9}$$

where $\boldsymbol{q}$ is the lowest endpoint in the support of $\boldsymbol{X}$ and $\mu(\cdot)$ is a measure called the exponent measure of G.

Let μ_* be the exponential measure of G_*; Equations (C.8) and (C.9) imply that for any $s > 0$

$$\mu_*(sB) = s^{-1}\mu_*(B) \tag{C.10}$$

for any Borel subset B; this is called the homogeneity property of G_*, and it can be meaningfully restated in polar coordinates in the following way: Let $\mathbb{H}_1 = \left\{\omega \in \mathbb{R}^d : \|\omega\|_1 = 1\right\}$ be the (n-1)-dimensional simplex in $\mathbb{R}^d$ and the mapping $T : \mathbb{R}^d \to (\boldsymbol{0}, \boldsymbol{\infty}) \times \mathbb{H}_1$ a pseudo-polar coordinates representation of any $z \in \mathbb{R}^d$ such that $T(\boldsymbol{z}) = (r, \boldsymbol{w})$, where $r = \|\boldsymbol{z}\|_1$ and $\boldsymbol{\omega} = \boldsymbol{z}/r$. Let us also define a measure S on $\Xi = \mathbb{H}_1 \cap [\boldsymbol{0}, \boldsymbol{\infty})$ as

$$S(B) = \mu_*\left(\{z \in [\boldsymbol{0}, \boldsymbol{\infty}) : \|\boldsymbol{z}\|_1 \geq \boldsymbol{1}, \boldsymbol{z}/\|\boldsymbol{z}\|_1 \in B\}\right) \tag{C.11}$$

$S(\cdot)$ is called the spectral measure of G_* and is uniquely determined by μ_*; the homogeneity property of equation C.10 implies

$$\mu_* \left(\{ \boldsymbol{z} \in [\boldsymbol{0}, \boldsymbol{\infty}) : \|\boldsymbol{z}\|_1 \geq r, \boldsymbol{z}/\|\boldsymbol{z}\|_1 \in B \} \right) = r^{-1} S(B) \tag{C.12}$$

The interpretation of this result is that the exponent measure μ_* can be factored as the product of a radial measure $r^{-2}dr$ and an angular measure S.

With some algebra, we can then express G_* as

$$G_*(\boldsymbol{z}) = \exp\left(- \int_\Xi \left(\bigvee_{j=1}^{d} \frac{\omega_j}{z_j} \right) S(\mathrm{d}\boldsymbol{\omega}) \right), \quad \boldsymbol{z} \in [\boldsymbol{0}, \boldsymbol{\infty}] \tag{C.13}$$

subject to

$$\int_\Xi \omega_j S(\mathrm{d}\boldsymbol{\omega}) = 1, \quad j = 1, \ldots, d \tag{C.14}$$

These constraints enforce standard Frechet margins as we originally assumed.

In view of these results, deriving a good model for the extremes of our data is a matter of accurately approximating the angular measure S; in practice, it is usually done by estimation of functionals that have a one to one correspondence with S, but that are more amenable for direct estimation, such as the stable tail dependence function (STDF), which has the additional property of being a bounded convex function, and is defined as follows

Definition 5 (stable tail dependence function)**.** *Let G_* be a max-stable distribution with Frechet margins and angular measure S, then its stable tail dependence function is defined as*

$$l(\boldsymbol{v}) = \int_\Xi \left(\bigvee_{j=1}^{d} v_j \omega_j \right) S(\mathrm{d}\boldsymbol{\omega}), \quad \boldsymbol{v} \in [\boldsymbol{0}, \boldsymbol{\infty}]$$

STDFs are convex and bounded by $||\boldsymbol{v}||_\infty \leq l(\boldsymbol{v}) \leq ||\boldsymbol{v}||_1$; they are directly related to G_* by

$$G_*(\boldsymbol{z}) = \exp(-l(1/z_1, ..., 1/z_d)) \tag{C.15}$$

STDFs also has the property that $l(s\boldsymbol{v}) = sl(\boldsymbol{v})$, so if we knew l for all values in the simplex $\mathbb{H}_1$ we could specify it completely; the Pickands dependence function, defined below for the two-dimensional case, takes advantage of this fact.

Definition 6 (Pickands dependence function)**.** *Let G_* be a max-stable distribution in $\mathbb{R}^2$ with Frechet margins and STDF l, then the Pickands dependence*

function is defined as

$$A(t) = l(1-t, t) \tag{C.16}$$

then for $(v_1, v_2) \in (\mathbb{R}^+)^2$ we have

$$l\left(v_1, v_2\right) = \left(v_1 + v_2\right) A\left(\frac{v_2}{v_1 + v_2}\right)$$

The Pickands dependence function and more generally the STDF are used in practice to indirectly estimate S, μ_* and G_* through (C.9)-(C.16). Once a model for G_* is available, inferences about the original data can be performed by applying the inverse transformation of (C.6).

There are many approaches to model $A(t)$ and $l(\boldsymbol{v})$, parametric and non-parametric; (Beirlant et al., 2006) contains detailed information on this topic. In our case, perhaps the most relevant model is the so-called logistic model, a parametric model that assumes exchangeability in the dependence of extreme values in all components of the data, and induces an STDF of the form

$$l(\boldsymbol{v}) = \left(\sum_{j=1}^{d} v_1^{1/\alpha}\right)^{\alpha}, \quad v_j \geq 0, \alpha \in [0, 1] \tag{C.17}$$

The smaller α is, the more dependent extreme values are across components: $\alpha = 0$ implies complete dependence, where all components are exactly equal, while $\alpha = 1$ implies statistical independence.

Yet another way of characterising G_* is by analysing the copula of the transformed data $\boldsymbol{z}_1, ..., \boldsymbol{z}_n$. The copula characterises the dependence structure between components and as such is not affected by component-wise transformations, so $\boldsymbol{X}$ and $\boldsymbol{Z}$ share the same copula. The definition of extreme value copulas, given below, is analogous to that of multivariate extreme value distributions

Definition 7 (Extreme value copula)**.** *C^* is an extreme value copula if for some copula C and every $\boldsymbol{u} \in [\boldsymbol{0}, \boldsymbol{1}]$*

$$C^*(\boldsymbol{u}) = \lim_{n \to \infty} C^n(u_1^{1/n}, ..., u_d^{1/n}) \tag{C.18}$$

and we say that C is in the domain of attraction of C^*.

Extreme value copulas are in a one-to-one relationship to STDFs through the following equality

$$C^*(\boldsymbol{u}) = \exp(-l(-\ln u_1, ..., -\ln u_d)) \tag{C.19}$$

One of the most popular parametric families of coupla models is the Archimedean family, and it turns out that, under mild smoothness assumptions, are in the domain of attraction of the Gumbel copula (Gudendorf and Segers, 2010), which induces a logistic STDF as defined in (C.17), so the logistic model is quite expressive provided the dependence structure is symmetric in the components.

Until now we have been assuming we know the marginal CDFs of $\boldsymbol{X}$ but often this is not the case, and they have to be approximated. Typically a semiparametric models is used for each component, with fitted generalised Pareto tails above a selected threshold, and an empirical CDF below it: let u_j be the threshold for the j-th component, $\tilde{F}_{\boldsymbol{X}_j}(\cdot)$ its empirical CDF, and σ_j, ξ_j its fitted tail GPD parameters, then its semiparametric approximation is given by

$$\hat{F}_{\boldsymbol{X}_j}(x) = \begin{cases} \tilde{F}_{X_j}(u_j) + \left(1 - \tilde{F}_{X_i}(u_j)\right)\left(1 - \left(1 + \xi_j \left(x - u_j\right)/\sigma_j\right)_+^{-1/\xi_j}\right), & x > u_j \\ \tilde{F}_{X_j}(x), \quad x \leq u_j \end{cases} \tag{C.20}$$

Thus, it is common in applications to use (C.20) when transforming marginal distributions in the data to and from Frechet or similar choices through (C.6)-(C.7).

C.2.3 Asymptotic dependence

Extreme value copulas, given by the limit (C.18), can take a wide variety of forms, and there is no closed form solution that encompasses all of them. One of the possible limits is in fact the independence copula. This would indicate that the dependence in extreme co-occurrences across components is weak or non-existent and vanishes at progressively more extreme levels, which can occur even though at non-extreme levels different components may be strongly associated; the multivariate normal distribution is a good example of this, as it is in the domain of attraction of the independent copula for any correlation ρ such that $-1 < \rho < 1$ (Sibuya, 1960). Because of this, when devising a multivariate extreme value model, care must be taken to make sure the chosen model reflects the data appropriately in this regard: misspecication of the extreme value copula may produce large errors in estimated risks. The concepts of asymptotic dependence and independence help us to address this issue: let $\boldsymbol{X} \in \mathbb{R}^2$ be a random vector with Frechet marginals. Define

$$\chi = \lim_{t \to \infty} \mathbb{P}(X_1 > t | X_2 > t) \tag{C.21}$$

We say that there is asymptotic dependence if $\chi > 0$, and asymptotic independence otherwise. A drawback of this metric is that it does not provide a way to characterise the rate at which the dependence fades away in the case of asymptotic independence. To address this, we can define the following statistic

$$\bar{\chi} = \lim_{p \to 1} \frac{2 \ln(1-p)}{\ln(\mathbb{P}(X_1 > F_{X_1}^{-1}(p), X_1 > F_{X_2}^{-1}(p))} \in [-1, 1] \tag{C.22}$$

We then have asymptotic dependence when $\bar{\chi} = 1$ and asymptotic independence otherwise. Moreover, values in $[-1, 1)$ are directly related to the rate with which the dependence fades away at extreme levels. However, accurate estimation of χ and $\bar{\chi}$ is not straightforward, and they are instead usually assessed graphically from direct empirical estimates; in this sense, yet another related statistic, the coefficient of tail dependence $\eta \in [0, 1]$, can be estimated by maximum likelihood through standard univariate EVT methods under the following additional assumption in the context of (C.21)

$$\mathbb{P}(X_1 > t, X_2 > t) \to \mathcal{L}(t) \times t^{-1/\eta} \tag{C.23}$$

where $\mathcal{L}(x)$ is a slowly varying function, that is, $\mathcal{L}(yz)/\mathcal{L}(y) \to 1$ as $y \to \infty$ for all $z > 0$. Intuitively, slowly varying functions become flat rapidly (in a precise sense) as $y \to \infty$ (note that $\mathcal{L}$ does not necessarily converge, e.g. $\mathcal{L}(z) = \ln(z)$). This class of functions arise in many important results from the theory of extreme values, and (C.23) has been shown to be valid for a broad range of conditions and models (Ledford and Tawn, 1997; Heffernan, 2000); here, η describes the type of dependence and $\mathcal{L}(y)$ its strength within the dependence type given by η. We have $\eta = 1$ whenever $\chi > 0$, and $0 \leq \eta < 1$ otherwise, thus characterising the presence of asymptotic dependence. Furthermore, we can estimate η by maximum likelihood as the shape parameter ξ in (C.5) from a sample of $W = \min\{Z_1, Z_2\}$. We also have the following identity

$$\eta = \frac{1}{2}(\bar{\chi} + 1) \tag{C.24}$$

In addition to these indices, there is abudant literature on statistical tests for asymptotic dependence/independence; (De Carvalho and Ramos, 2012) provides a good overview of this topic.

C.2.4 Multivariate Generalised Pareto distributions

Just as (C.5) for the univariate case, asymptotic results for the distribution of exceedances above a suitable threshold can be derived in the multivariate case. Let the operator $\boldsymbol{x} \not\leq \boldsymbol{y}$ denote an exceedance of $\boldsymbol{x}$ above $\boldsymbol{y}$ in at least one component, i.e., $x_i > y_i$ for some $i = 1, ..., d$. Let $\boldsymbol{x}_1, ..., \boldsymbol{x}_n \in \mathbb{R}^d$ be a sample from $F_{\boldsymbol{X}} \in \mathcal{D}(G)$, i.e., $G(\boldsymbol{x})$ is the limiting multivariate extreme value distribution for normalised maxima of $\boldsymbol{X}$. Then it can be shown that there are normalising constants $\{\boldsymbol{a}_n\}, \{\boldsymbol{b}_n\}$ such that

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\boldsymbol{X} - \boldsymbol{b}_n}{\boldsymbol{a}_n} \leq \boldsymbol{x} | \boldsymbol{X} \not\leq \boldsymbol{b}_n\right) = \frac{\ln(G(\boldsymbol{x})) - \ln(G(\boldsymbol{x} \wedge \boldsymbol{0}))}{-\ln(G(\boldsymbol{0}))}, \quad \boldsymbol{x} \not\leq \boldsymbol{0} \tag{C.25}$$

where $\boldsymbol{x} \wedge \boldsymbol{0}$ is a component-wise minimum operator, and $G(\boldsymbol{0}) > 0$ without loss of generality (Beirlant et al., 2006). Distributions of the form (C.25), where G is a max-stable distribution, are called multivariate Generalised Pareto distributions.

Definition 8 (Multivariate Generalised Pareto distributions)**.** *Let G be a max-stable distribution, and assume without loss of generality that $G(\boldsymbol{0}) > 0$, then the associated multivariate Generalised Pareto distribution is given by*

$$F_{\boldsymbol{Z}}(\boldsymbol{z}) = \frac{\ln(G(\boldsymbol{z})) - \ln(G(\boldsymbol{z} \wedge \boldsymbol{0}))}{-\ln(G(\boldsymbol{0})}, \quad \boldsymbol{z} \not\leq \boldsymbol{0} \tag{C.26}$$

The one-to-one relationship between multivariate GPD and extreme value distributions is clear from (C.26); furthermore, for practical applications, an argument similar to (C.3) justifies their use as multivariate exceedance models.

Let $\boldsymbol{W}$ be a multivariate GPD, such that $F_{\boldsymbol{W}}$ is given by (C.26). Note the support of $\boldsymbol{W}$ has an inverted L-shape (or its multidimensional analogous), and so it is not a product space; note also that marginal distributions $W_j,\ j = 1, ..., d$ are not univariate GPD distributions, however, marginal conditional exceedances of the type $W_j | W_j > 0,\ j = 1, ..., d$ are (Rootzén and Tajvidi, 2006); furthermore, conditional exceedances of $\boldsymbol{W}$ only shift its location parameter, i.e., if $\boldsymbol{c} > \boldsymbol{0}$ and $\boldsymbol{V} = \boldsymbol{W} \mid \boldsymbol{W} \not\leq \boldsymbol{c}$, then $F_{\boldsymbol{V}}(\boldsymbol{v}) = F_{\boldsymbol{W}}(\boldsymbol{v} - \boldsymbol{c})$.

C.2.5 Conditional extremes

Parametric models of multivariate extreme value distributions make the implicit assumption that extreme co-ocurrences across components are asymptotically dependent. This is because limiting results for asymptotically independent distributions are either full independence, which is not informative, or degenerate

distributions; an example for the latter is (C.26), which places all of its probability mass on regions with an infinite negative component if the underlying distribution $F_{\boldsymbol{W}}$ is asymptotically independent. In (Heffernan and Tawn, 2004), this problem is addressed by developing a framework for the analysis of conditional extremes instead. This is done under a similar picture to usual extreme value analysis where normalised sample maxima is assumed to converge in distribution to a non-degenerate limit, but in this case this is only assumed to hold when conditioning to an exceedance in some component. To make this clearer, let $\boldsymbol{Y}$ be a vector with Gumbel marginal distributions without loss of generality; let $\boldsymbol{Y}_{-i}$ be the sub vector of all components of $\boldsymbol{Y}$ except the i-th entry, and let us also assume that for every component $i = 1, ..., d$ there exist vector functions $\boldsymbol{a}_i(y_i) \in (\mathbb{R}^+)^{(d-1)}, \boldsymbol{b}_i(y_i) \in \mathbb{R}^{(d-1)}$ such that

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\boldsymbol{Y}_{-i} - \boldsymbol{b}_i(y_i)}{\boldsymbol{a}_i(y_i)} \leq \boldsymbol{z} \mid y_i = n\right) = G_{|i}(\boldsymbol{z}) \tag{C.27}$$

where $G_{|i}(\boldsymbol{z})$is nondegenerate. The authors develop a full semi-parametric methodology for estimating $G_{|i}$ together with tests for goodness-of-fit. Let $\mathbf{Z}_{|i} = (\boldsymbol{Y}_{-i} - \boldsymbol{b}_i(y_i))/\boldsymbol{a}_i(y_i)$, then a crucial observation is the fact that

$$\lim_{u_i\to\infty} \mathbb{P}\left(\mathbf{Z}_{|i} \leqslant \mathbf{z}, Y_i - u_i = y \mid y_i > u_i\right) = G_{|i}(\boldsymbol{z}) \exp(-y) \tag{C.28}$$

This equation tells us that $\boldsymbol{Z}_{|i}$ is independent from the exceedance $Y_i - u_i$ in the limit, and that the limiting distribution for the latter is exponential; these two facts allows the simulation of joint samples from the model once $G_{|i}$ has been estimated by choosing a large enough threshold u_i, then simulating $\boldsymbol{Z}_{|i}$ independently using the fitted semi-parametric residual model. The authors also provide closed-form results for $G_{|i}$ for some parametric models, one of which is the logistic model, whose conditional extreme distribution is given by

$$G_{|i}\left(\mathbf{z}\right) = \left\{1 + \sum_{j\neq i} \exp\left(-\frac{z_i}{\alpha}\right)\right\}^{\alpha-1} \tag{C.29}$$

which resembles a multivariate version of a type I generalised logistic distribution. This methodology is amenable for simulating the data distribution conditioned to an exceedance in a given component. However, consistency issues arise when simulating data conditioned to an exceedance in *any* component. It is argued in (Heffernan and Tawn, 2004) that this is only a minor issue, as the problematic region (i.e., the region of joint exceedances) has a small probability mass compared to other exceedance regions; some of these consistency issues were subsequently

addressed in (Keef et al., 2013; Liu and Tawn, 2014).

Appendix D

Chapter 5 appendix

D.1 Experiment data

Peak Season	**ACS peak (MW)**	**Start**	**End**
2005/6	60,440	30-Oct-05	25-Mar-06
2006/7	59,948	29-Oct-06	24-Mar-07
2007/8	59,762	28-Oct-07	22-Mar-08
2008/9	57,474	26-Oct-08	21-Mar-09
2009/10	57,290	25-Oct-09	20-Mar-10
2010/11	57,675	31-Oct-10	26-Mar-11
2011/12	55,954	30-Oct-11	24-Mar-12
2012/13	55,309	29-Oct-12	24-Mar-13
2013/14	54,422	27-Oct-13	22-Mar-14
2014/15	54,072	26-Oct-14	21-Mar-15
2015/16	52,164	26-Oct-15	20-Mar-16
2016/17	51,525	31-Oct-16	26-Mar-17

Table D.1: Official peak season data used by National Grid for the period 2005-2016

Scenario	Onshore	Offshore	Total (GW)
1	0.0	0.0	0.0
2	14.1	10.5	24.6
3	18.2	22.5	40.7
4	26.5	44.5	71.0

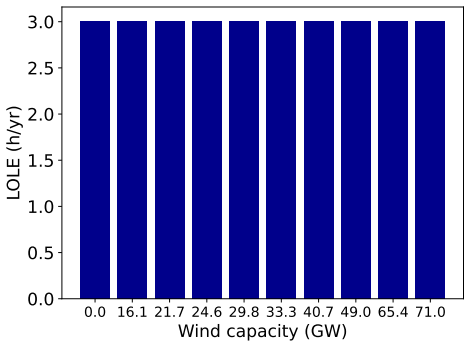
Table D.2: Scenarios of installed wind capacity used in experiments from Section 5.7

Scenario	Onshore	Offshore	Total (GW)
1	0.0	0.0	0.0
2	10.8	5.3	16.1
3	13.5	8.2	21.7
4	14.1	10.5	24.6
5	15.0	14.8	29.8
6	16.0	17.3	33.3
7	18.2	22.5	40.7
8	21.9	27.1	49.0
9	24.8	40.5	65.3
10	26.5	44.5	71.0

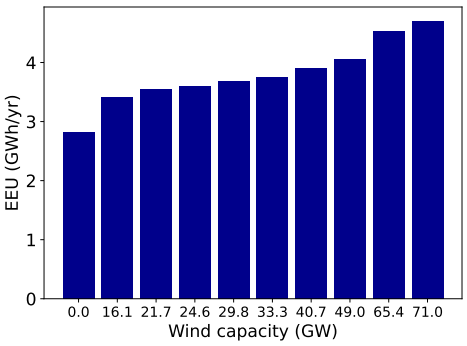
Table D.3: Scenarios of installed wind capacity used in experiments from Section 5.6

D.2 Results for different experimental configurations

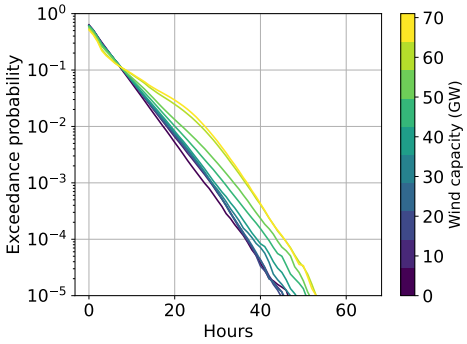
Below, figures for the full results for all experimental configurations (Table 5.2) from Section 5.6 are shown. For every configuration, both LOLE and EEU figures are shown at the top and display values for all of the tested wind capacities; recall that some scenarios assume a fixed LOLE, while others assume a fixed EEU. Beyond summary statistics, figures show the survival functions (in log scale) for the calculated risk metrics which include EU, LOLD, LOLC, number of days with shortfalls, EU conditioned to days with shortfall and LOLD conditioned to days with shortfall. The Y axis is the same across all of these figures for ease of comparison.



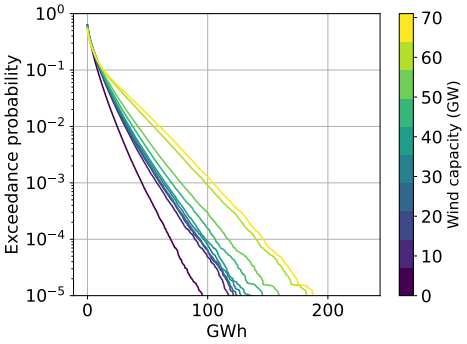
(a) LOLE



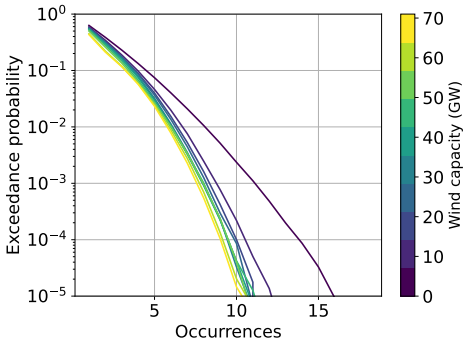
(b) EEU



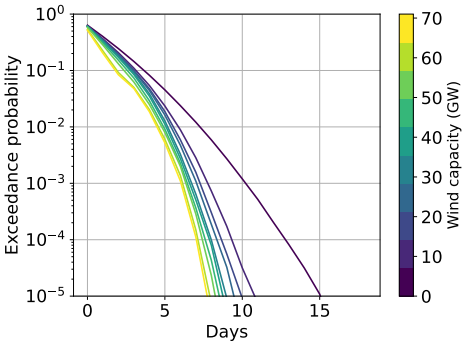
(c) Loss of load duration



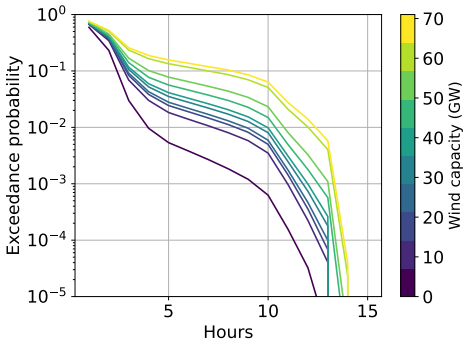
(d) Energy unserved



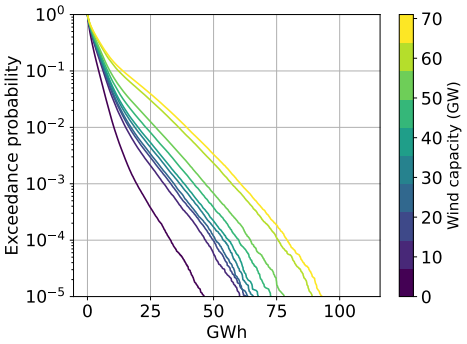
(e) Loss of load count



(f) Number of days of shortfall

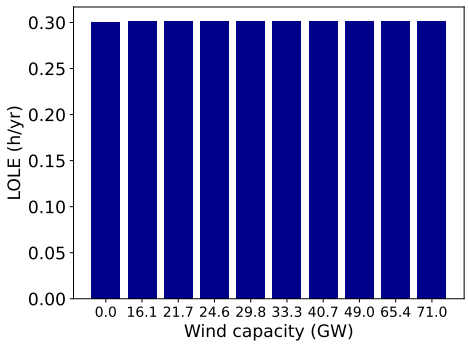


(g) LOLD within days of shortfall

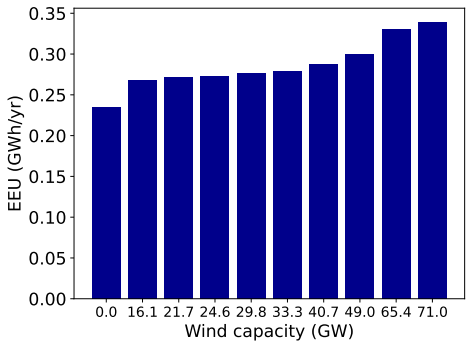


(h) EU within days of shortfall

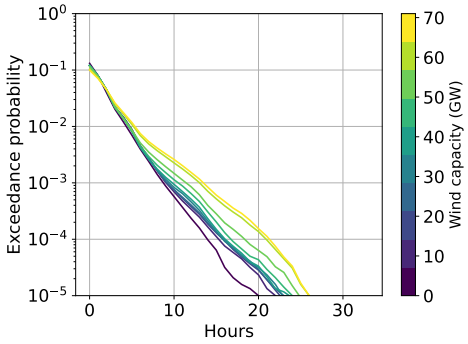
Figure D.1: Standard scenario



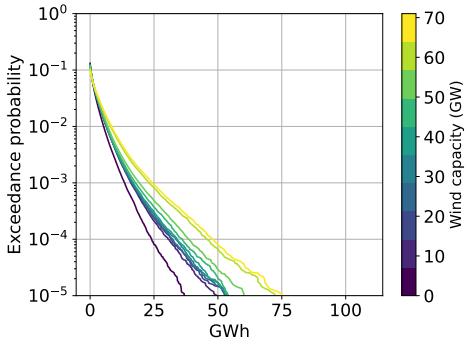
(a) LOLE



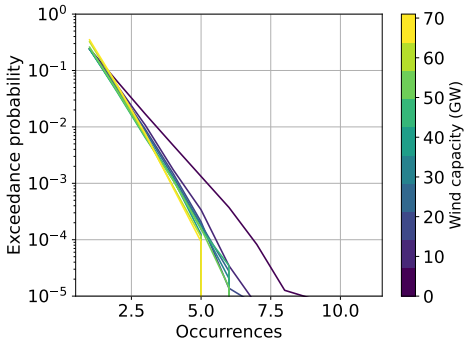
(b) EEU



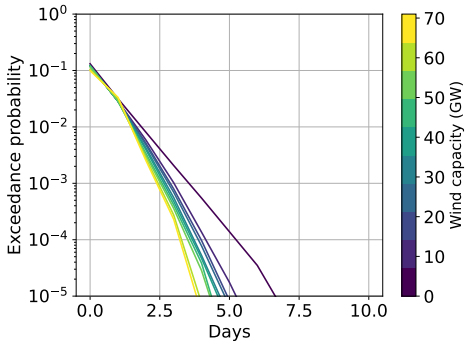
(c) Loss of load duration



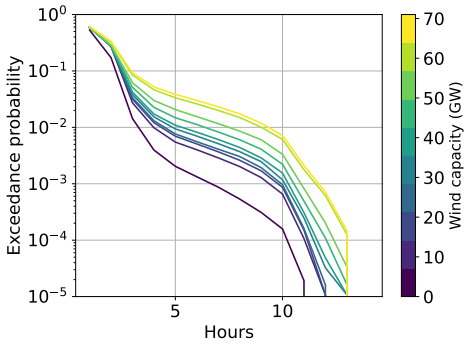
(d) Energy unserved



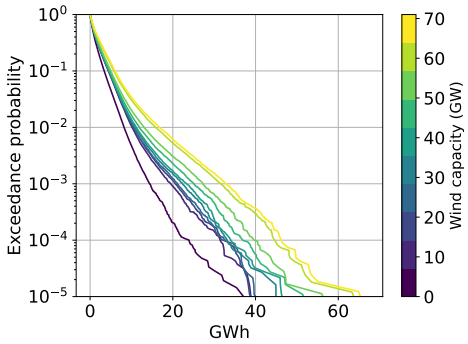
(e) Loss of load count



(f) Number of days of shortfall

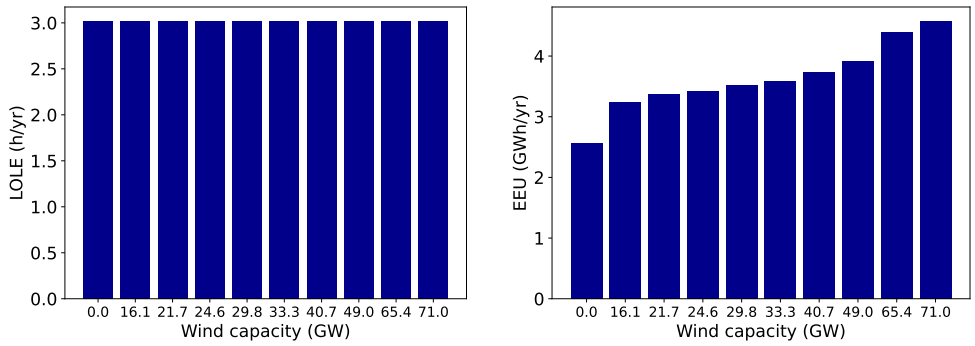


(g) LOLD within days of shortfall



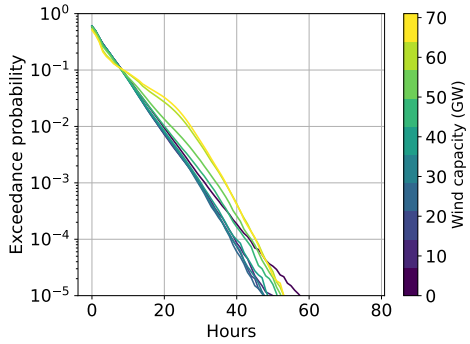
(h) EU within days of shortfall

Figure D.2: Low LOLE scenario

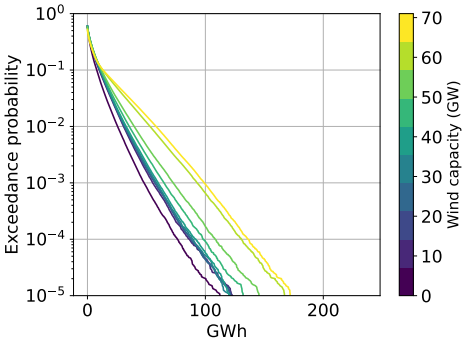


(a) LOLE

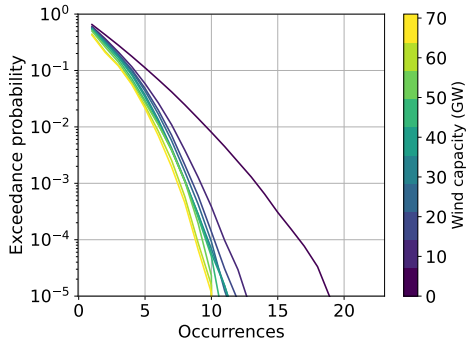
(b) EEU



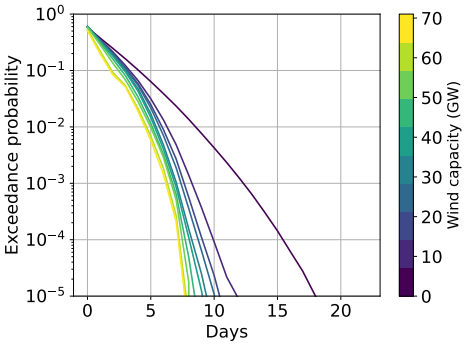
(c) Loss of load duration



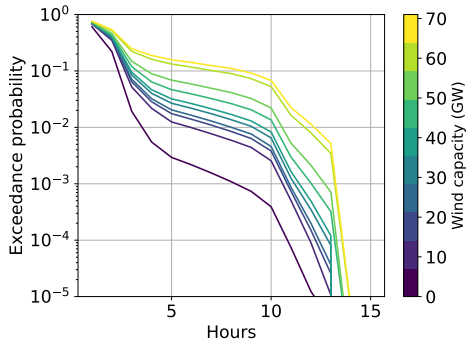
(d) Energy unserved



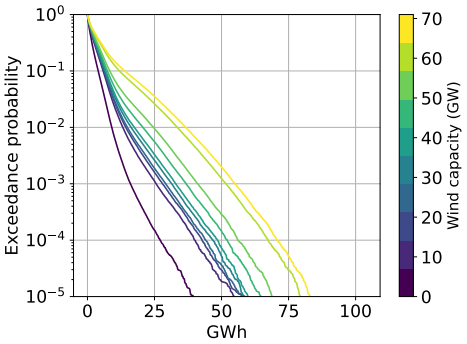
(e) Loss of load count



(f) Number of days of shortfall

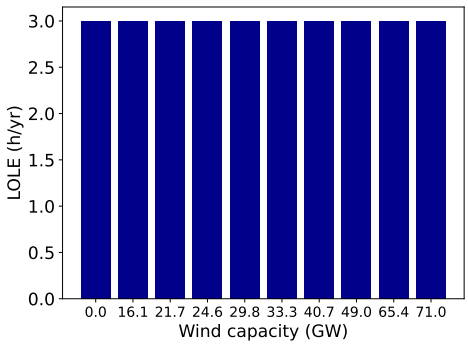


(g) LOLD within days of shortfall

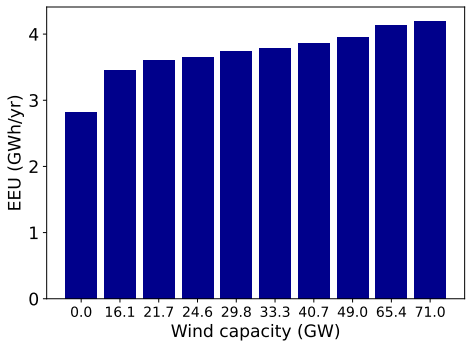


(h) EU within days of shortfall

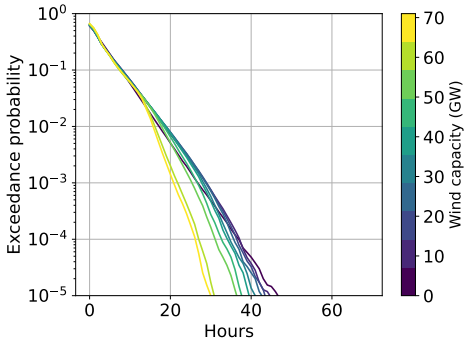
Figure D.3: Scenario of ACG model with data from (Murphy et al., 2018)



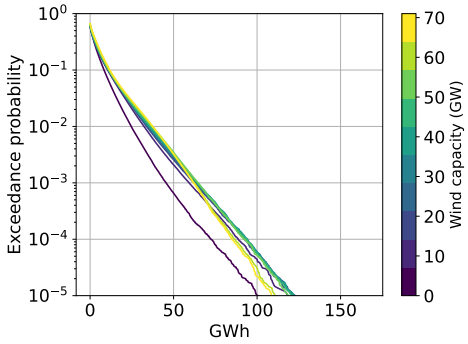
(a) LOLE



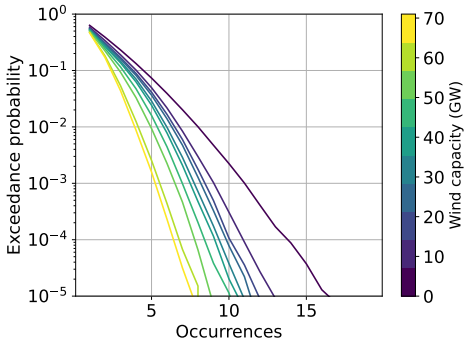
(b) EEU



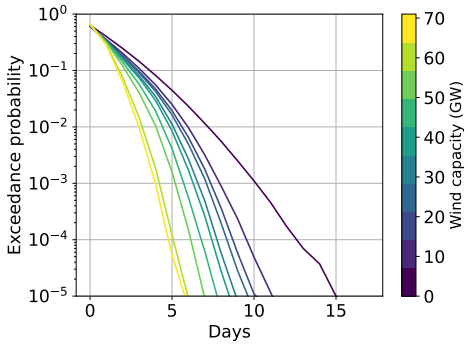
(c) Loss of load duration



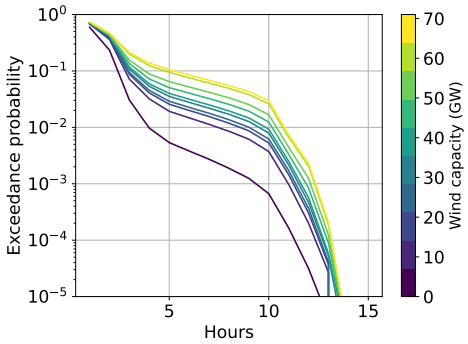
(d) Energy unserved



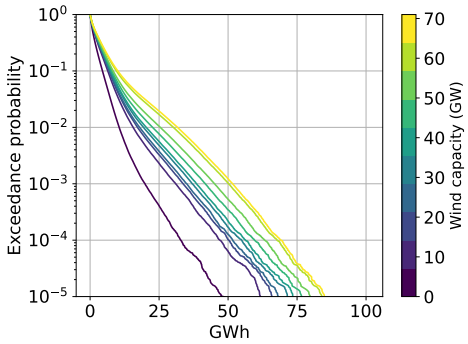
(e) Loss of load count



(f) Number of days of shortfall

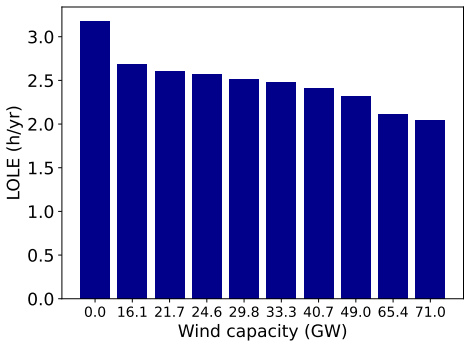


(g) LOLD within days of shortfall

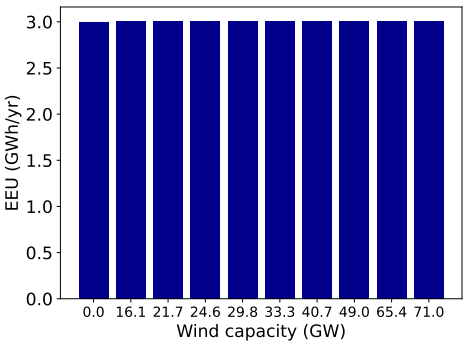


(h) EU within days of shortfall

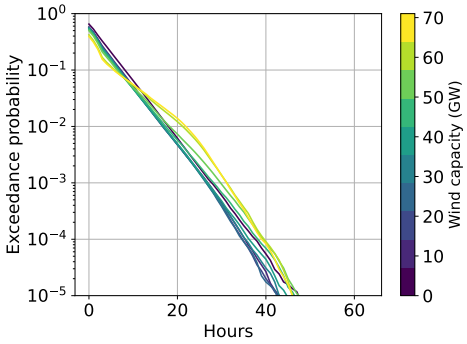
Figure D.4: Scenario without low-wind year of 2005



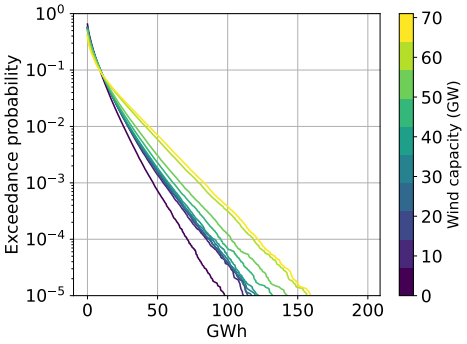
(a) LOLE



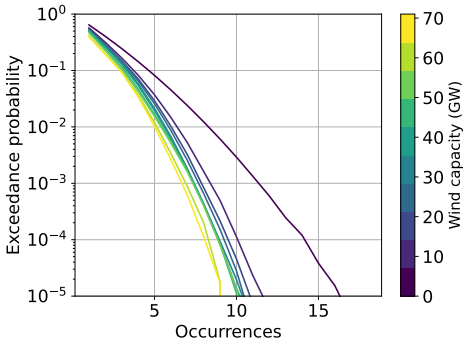
(b) EEU



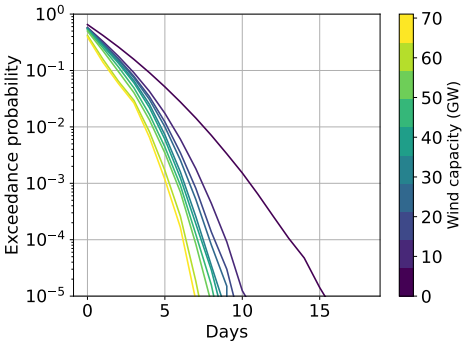
(c) Loss of load duration



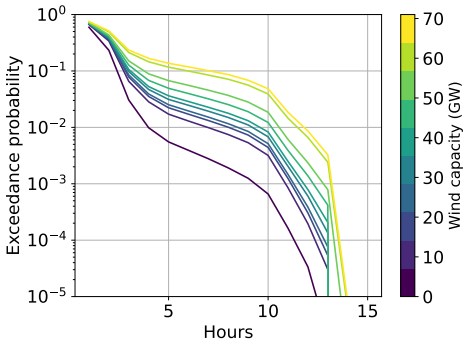
(d) Energy unserved



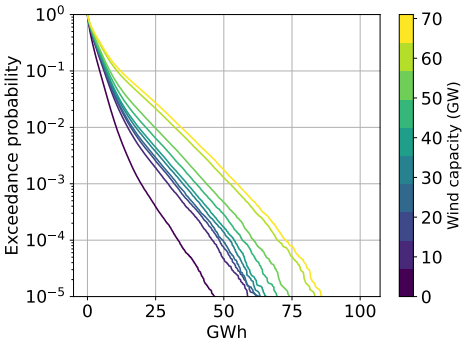
(e) Loss of load count



(f) Number of days of shortfall



(g) LOLD within days of shortfall

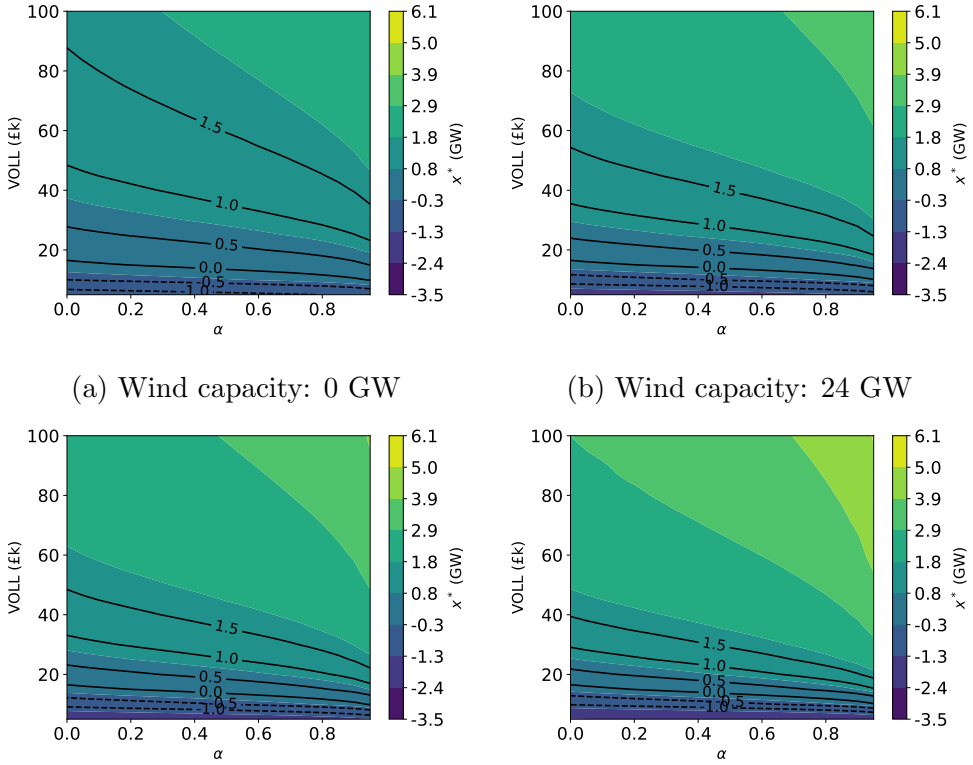


(h) EU within days of shortfall

Figure D.5: Scenario with fixed EEU = 3 GWh/year

D.3 Results for raw VOLL values

The following figures show the results from Section 5.7.3 expressed in terms of raw VOLL values.



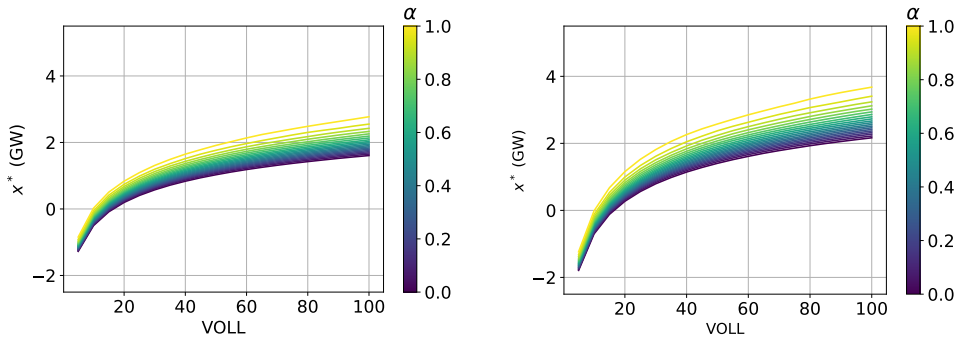
(a) Wind capacity: 0 GW

(b) Wind capacity: 24 GW

(c) Wind capacity: 41 GW

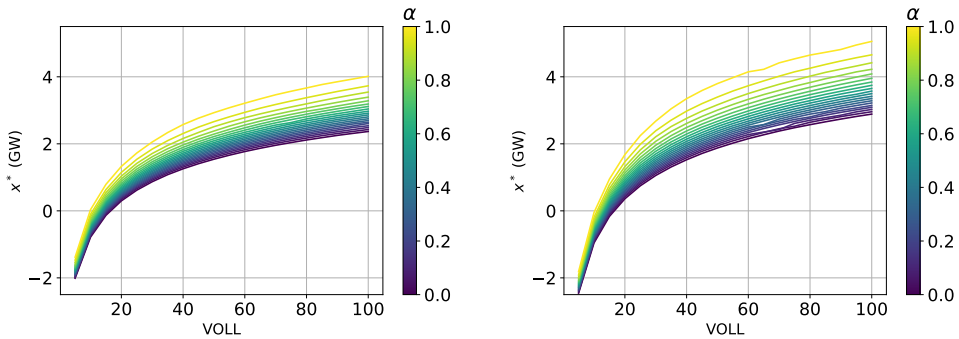
(d) Wind capacity: 71 GW

Figure D.6: Contour lines for the optimal procured capacity x^* as a function of the risk aversion parameter α (X axis) and the VOLL parameter (Y axis), for each wind capacity scenario. Dotted lines denote negative contour values.



(a) Wind capacity: 0 GW

(b) Wind capacity: 24 GW



(c) Wind capacity: 41 GW

(d) Wind capacity: 71 GW

Figure D.7: Optimal procured capacity x^* (Y axis) as a function of VOLL (X axis) for each of the wind capacity scenarios.

Bibliography

A. Pina, C. Silva, P. F. (2011). Modeling hourly electricity dynamics for policy making in long-term scenarios. *Energy Policy*, 39(9):4692–4702.

Antweiler, W. and Muesgens, F. (2021). On the long-term merit order effect of renewable energies. *Energy Economics*, 99:105275.

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent Measures of Risk. *Mathematical Finance*, 9(3):203–228.

Baldwin, C. J., Billings, J. E., Gaver, D. P., and Hoffman, C. H. (1959a). Mathematical Models for Use in the Simulation of Power Generation Outages II-Power System Forced-Outage Distributions. *Transactions of the American Institute of Electrical Engineers. Part III: Power Apparatus and Systems*, 78(4):1258–1267.

Baldwin, C. J., Gaver, D. P., and Hoffman, C. H. (1959b). Mathematical Models for Use in the Simulation of Power Generation Outages I-Fundamental Considerations. *Transactions of the American Institute of Electrical Engineers. Part III: Power Apparatus and Systems*, 78(4):1251–1258.

Baldwin, C. J., Gaver, D. P., Hoffman, C. H., and Rose, J. A. (1959c). Mathematical Models for Use in the Simulation of Power Generation Outages Part III. Models for a Large Interconnection. *Transactions of the American Institute of Electrical Engineers. Part III: Power Apparatus and Systems*, 78(4):1645–1649.

Ballester, C. and Furió, D. (2015). Effects of renewables on the stylized facts of electricity prices. *Renewable and Sustainable Energy Reviews*, 52:1596–1609.

Baringa Ltd (2016). Baringa SEM PLEXOS Forecast Model for 2016-17. Technical report, Baringa Ltd. available at https://www.semcommittee.com/news-centre/baringa-sem-plexos-forecast-model-2016-17.

Basaran Filik, U., Gerek, O. N., and Kurban, M. (2009). Hourly Forecasting of Long Term Electric Energy Demand Using a Novel Modeling Approach. In

2009 Fourth International Conference on Innovative Computing, Information and Control (ICICIC), pages 115–118.

Beirlant, J., Goegebeur, Y., and Segers, J. (2006). *Statistics of Extremes*. John Wiley and Sons.

BEIS (2016). Energy consumption in the UK. Technical report, Department for Business, Energy and Industrial Strategy.

Bell, D. E. (1982). Regret in Decision Making under Uncertainty. *Operations Research*, 30(5):961–981.

Benner, P. (1934). The use of the theory of probability to determine spare capacity. *General Electric Review*, 37(7):345–348.

Billinton, R. and Allan, R. (1984). Power-system reliability in perspective. *Electronics and Power*, 30(3):231–236.

Billinton, R. and Allan, R. N. (1985). Reliability evaluation of power systems. *Quality and Reliability Engineering International*, 1(2):141–141.

Billinton, R. and Jonnavithula, A. (1997). Composite system adequacy assessment using sequential Monte Carlo simulation with variance reduction techniques. *IEE Proceedings - Generation, Transmission and Distribution*, 144:1–6(5).

Billinton, R., Wacker, G., and Wojczynski, E. (1983). Comprehensive Bibliography On Electrical Service Interruption Costs. *IEEE Transactions on Power Apparatus and Systems*, PAS-102(6):1831–1837.

Boßmann, T. and Staffell, I. (2015). The shape of future electricity demand: Exploring load curves in 2050s Germany and Britain. *Energy*, 90:1317–1333.

Calabrese, G. (1947). Generating Reserve Capacity Determined by the Probability Method. *Transactions of the American Institute of Electrical Engineers*, 66(1):1439–1450.

Castellanos, M. and Cabras, S. (2007). A default Bayesian procedure for the generalized Pareto distribution. *Journal of Statistical Planning and Inference*, 2.

Chen, M.-H. (2005). Bayesian Computation: From Posterior Densities to Bayes Factors, Marginal Likelihoods, and Posterior Model Probabilities. In Dey, D.

and Rao, C. R., editors, *Bayesian Thinking, Modeling and Computation*, chapter 15, pages 449–450. Elsevier Science.

Clancy, D., Gross, G., and Wu, F. (1983). Probabilitic flows for reliability evaluation of multiarea power system interconnections. *International Journal of Electrical Power And Energy Systems*, 5(2):101–114.

Coles, S. (2013). *An Introduction to Statistical Modeling of Extreme Values*. Springer London.

Cook, V. M., Steinberg, M. J., Galloway, C. D., and Wood, A. J. (1963). Determination or reserve requirements for two interconnected systems. *IEEE Trans. Power Apparatus and Systems*, 82(65):18–33.

Council of European Union (2013). REGULATION (EU) No 347/2013 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL . https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:115:0039:0075:en:PDF.

D'Amico, G., Petroni, F., and Prattico, F. (2015). Wind speed prediction for wind farm applications by Extreme Value Theory and Copulas. *Journal of Wind Engineering and Industrial Aerodynamics*, 145:229–236.

De Carvalho, M. and Ramos, A. (2012). Bivariate Extreme Statistics, II. *Revstat - Statistical Journal*, 10:81–104.

DECC (2013a). ELECTRICITY GENERATION COST MODEL - 2013 UPDATE OF NON-RENEWABLE TECHNOLOGIES. Technical report, Department of Energy and Climate Change. Available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/223634/2013_Update_of_Non-Renewable_Technologies_FINAL.pdf.

DECC (2013b). Electricity Market Reform: Consultation on Proposals for Implementation. Technical report, Department of Energy and Climate Change. Available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/255254/emr_consultation_implementation_proposals.pdf.

DECC (2013c). Electricity Market Reform Delivery Plan. Technical report, Department of Energy and Climate Change. Available at https://assets.publishing.service.gov.uk/government/uploads/

system/uploads/attachment_data/file/268221/181213_2013_EMR_Delivery_Plan_FINAL.pdf.

DECC (2013d). EMR Consultation Annex C: Reliability Standard Methodology. Technical report, Department of Energy and Climate Change.

Decker, M., Brunke, M., Wang, Z., Sakaguchi, K., Zeng, X., and Bosilovich, M. (2012). Evaluation of the reanalysis products from gsfc, ncep, and ecmwf using flux tower observations. *Journal of Climate*, 25(6):1916–1944.

Dent, C. J. and Zachary, S. (2014). Further results on the probability theory of capacity value of additional generation. In *2014 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, pages 1–6.

Dias, G. and da Costa, L. (2018). Using cvar for adequacy assessments and in resource expansion models. Working Group on LOLE Best Practices.

Dickey, J. and Lientz, B. (1970). The Weighted Likelihood Ratio, Sharp Hypotheses about Chances, the Order of a Markov Chain. *Annals of Mathematical Statistics*, 41.

Ecorys (2022). The value of lost load for electricity in the Netherlands. Technical report, Ecorys. Available at https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwjY4ouQ5fz6AhXUBEQIHYiUBSQQFnoECA4QAQ&url=https%3A%2F%2Fenergeia-binary-external-prod.imgix.net%2FNGDmry3p7sTJMqLJehNj4jwcsHg.pdf%3Fdl%3DThe%2Bvalue%2Bof%2Blost%2Bload%2Bfor%2Belectricity%2Bin%2Bthe%2BNetherlands.pdf&usg=AOvVaw1RzDbL4EDBqtwa4hXyfHW6.

Edwards, G., Sheehy, S., Dent, C. J., and Troffaes, M. C. (2017). Assessing the contribution of nightly rechargeable grid-scale storage to generation capacity adequacy. *Sustainable Energy, Grids and Networks*, 12:69–81.

Elliott, D. F. (1987). Chapter 7 - fast fourier transforms. In Elliott, D. F., editor, *Handbook of Digital Signal Processing*, pages 527–631. Academic Press, San Diego.

Embrechts, P., Resnick, S., and Samorodnitsky, G. (1999). Extreme value theory as a risk management tool. soa seminar: Integrated approaches to risk measurement in the financial services industry (atlanta, ga, 1997). *North American Actuarial Journal*, 3.

Engelke, S. and Hitz, A. S. (2018). Graphical models for extremes. *arXiv preprint arXiv:1812.01734*.

Engie (2016). Understanding the capacity market. Technical report, Engie.

EU (2018). 2nd ENTSO-E Guideline For Cost Benefit Analysis of Grid Development Projects. Technical report, European Comission. Available at https://eepublicdownloads.entsoe.eu/clean-documents/tyndp-documents/Cost%20Benefit%20Analysis/2018-10-11-tyndp-cba-20.pdf.

European Comission (2016). Final Report of the Sector Inquiry on Capacity Mechanisms. Technical report, European Comission. available at https://ec.europa.eu/energy/sites/ener/files/documents/com2016752.e\n_.pdf.

Fan, S. and Hyndman, R. J. (2012). Forecasting electricity demand in australian national electricity market. In *2012 IEEE Power and Energy Society General Meeting*, pages 1–4.

Ferrari, J. (2021). *Electric Utility Resource Planning*. Elsevier.

Fishburn, P. (1982). *The Foundations of Expected Utility*. Theory & Decision Librari Springer.

Fishman, G. S. (1996). *Generating Samples*, pages 145–254. Springer New York, New York, NY.

Galloway, C. D., Garver, L. L., Ringlee, R. J., and Wood, A. J. (1969). Frequency and Duration Methods for Power System Reliability Calculations Part III: Generation System Planning. *IEEE Transactions on Power Apparatus and Systems*, PAS-88(8):1216–1223.

Gelaro, R., McCarty, W., Suárez, M., Todling, R., Molod, A., Takacs, L., Randles, C., Darmenov, A., Bosilovich, M., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., Da Silva, A., Gu, W., and Zhao, B. (2017). The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *Journal of Climate*, 30.

Gorman, W. (2022). The quest to quantify the value of lost load: A critical review of the economics of power outages. *The Electricity Journal*, 35(8):107187.

Grechuk, B. and Zabarankin, M. (2014). Risk averse decision making under catastrophic risk. *European Journal of Operational Research*, 239(1):166–176.

Greenwood, D. M., Deakin, M., Sarantakos, I., Brayshaw, D. J., and Bloomfield, H. C. (2022). Capacity value of interconnectors for resource adequacy assessment in multi-region systems. In *2022 17th International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, pages 1–6.

Grigg, C., Wong, P., Albrecht, P., Allan, R., Bhavaraju, M., Billinton, R., Chen, Q., Fong, C., Haddad, S., Kuruganty, S., Li, W., Mukerji, R., Patton, D., Rau, N., Reppen, D., Schneider, A., Shahidehpour, M., and Singh, C. (1999). The IEEE Reliability Test System-1996. A report prepared by the Reliability Test System Task Force of the Application of Probability Methods Subcommittee. *IEEE Transactions on Power Systems*, 14(3):1010–1020.

Gudendorf, G. and Segers, J. (2010). Extreme-value copulas. In Jaworski, P., Durante, F., Härdle, W. K., and Rychlik, T., editors, *Copula Theory and Its Applications*, pages 127–145, Berlin, Heidelberg. Springer Berlin Heidelberg.

Hagspiel, S., Knaut, A., and Peter, J. (2018). Reliability in Multi-regional Power Systems: Capacity Adequacy and the Role of Interconnectors. *The Energy Journal*, 0(Number 5).

Hagspiel, S., Papaemannouil, A., Schmid, M., and Andersson, G. (2012). Copula-based modeling of stochastic wind power in Europe and implications for the Swiss power grid. *Applied Energy*, 96:33–44. Smart Grids.

Hall, J. D., Ringlee, R. J., and Wood, A. J. (1968). Frequency and Duration Methods for Power System Reliability Calculations: I - Generation System Model. *IEEE Transactions on Power Apparatus and Systems*, PAS-87(9):1787–1796.

Hall, P. and Tajvidi, N. (2000). Distribution and dependence-function estimation for bivariate extreme-value distributions. *Bernoulli*, 6(5):835 – 844.

Hartman, R. S., Doane, M. J., and Woo, C.-K. (1991). Consumer rationality and the status quo. *The Quarterly Journal of Economics*, 106(1):141–162.

Hausman, J. (2012). Contingent valuation: From dubious to hopeless. *Journal of Economic Perspectives*, 26(4):43–56.

Heffernan, J. (2000). A Directory of Coefficients of Tail Dependence. *Extremes*, 3:279–290.

Heffernan, J. E. and Tawn, J. A. (2004). A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(3):497–546.

Hu and Bo (2013). Copula-ARMA Model for Multivariate Wind Speed and Its Applications in Reliability Assessment of Generating Systems . *Journal of Electrical Engineering and Technology*, 8(3):421–427.

IEEE (2019a). IEEE Standard Terms for Reporting and Analyzing Outage Occurrences and Outage States of Electrical Transmission Facilities. *IEEE Std 859-2018 (Revision of IEEE Std 859-1987)*, pages 1–24.

IEEE (2019b). Ieee standard terms for reporting and analyzing outage occurrences and outage states of electrical transmission facilities. *IEEE Std 859-2018 (Revision of IEEE Std 859-1987)*, pages 1–24.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291.

Kaufman, A. and Daly (1978). The Cost of an Urban Blackout, The Consolidated Edison Blackout, July 13-14, 1977. *The Library of Congress Congressional Research Service, Washington, D.C.*

Keane, A., Milligan, M., Dent, C., Hasche, B., D'Annunzio, C., Dragoon, K., Holttinen, H., Samaan, N., Soder, L., and O'Malley, M. (2011). Capacity value of wind power. IEEE Trans Power Syst. *Power Systems, IEEE Transactions on*, 26:564 – 572.

Keef, C., Papastathopoulos, I., and Tawn, J. A. (2013). Estimation of the conditional distribution of a multivariate variable given that one of its components is large: Additional constraints for the heffernan and tawn model. *Journal of Multivariate Analysis*, 115:396–404.

Koreneff, G., Ruska, M., Kiviluoma, J., Shemeikka, J., Lemström, B., Alanen, R., and Koljonen, T. (2009). Future development trends in electricity demand. Technical report, VTT Technical Research Centre of Finland. Available at https://cris.vtt.fi/en/publications/future-development-trends-in-electricity-demand.

Lago-Gonzalez and Singh, C. (1990). The extended decomposition-simulation approach for multi-area reliability calculations. *IEEE Transactions on Power Systems*, 5(3):1024–1031.

Ledford, A. W. and Tawn, J. A. (1997). Modelling Dependence Within Joint Tail Regions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 59(2):475–499.

Li, W. (1994). *Reliability Assessment of Electric Power Systems Using Monte Carlo Methods*. Language of science. Springer US.

Liu, Y. and Tawn, J. (2014). Self-consistent estimation of conditional multivariate extreme value distributions. *Journal of Multivariate Analysis*, 127:19–35.

Loane, E. S. and Watchorn, C. W. (1947). Probability Methods Applied to Generating Capacity Problems of a Combined Hydro and Steam System. *Transactions of the American Institute of Electrical Engineers*, 66(1):1645–1657.

Loomes, G. and Sugden, R. (1982). Regret Theory: An Alternative Theory of Rational Choice Under Uncertainty. *The Economic Journal*, 92(368):805–824.

Lyman, W. J. (1947). Calculating Probability of Generating Capacity Outages. *Transactions of the American Institute of Electrical Engineers*, 66(1):1471–1477.

M. Grams, C., Beerli, R., Pfenninger, S., Staffell, I., and Wernli, H. (2017). Balancing Europe's Wind-Power Output Through Spatial Deployment Informed by Weather Regimes. *Nature Climate Change*, 7.

Mays, J., Morton, D., and O'Neill, R. (2019). Asymmetric risk and fuel neutrality in electricity capacity markets. *Nature Energy*, 4.

McFadden, D. (1986). The choice theory approach to market research. *Marketing Science*, 5(4):275–297.

Meyer, J. (2014). Chapter 3 - The Theory of Risk and Risk Aversion. In Machina, M. and Viscusi, K., editors, *Handbook of the Economics of Risk and Uncertainty*, volume 1 of *Handbook of the Economics of Risk and Uncertainty*, pages 99–133. North-Holland.

Morales González, J., Conejo, A., Madsen, H., Pinson, P., and Zugno, M. (2014). *Integrating Renewables in Electricity Markets: Operational Problems*. International Series in Operations Research and Management Science. Springer.

Murphy, S., Apt, J., Moura, J., and Sowell, F. (2018). Resource adequacy risks to the bulk power system in North America. *Applied Energy*, 212:1360–1376.

Murphy, S., Lavin, L., and Apt, J. (2020). Resource adequacy implications of temperature-dependent electric generator availability. *Applied Energy*, 262:114424.

Myers, D. (1978). The Economic Effects to a Metropolitan Area of Power Outage Resulting from an Earthquake. *Earthquake Engineering Systems Inc., San Francisco*.

National Grid (2017a). Average Cold Spell Methodology. Technical report, NGESO. available at https://www.emrdeliverybody.com/Lists/Latest%20News/Attachments/189/SC4L12%20ACS%20Methodology.pdf.

National Grid (2017b). Security of Supply - International Review of Standards and Implementation. Technical report, IEEE Working Group of LOLE Best Practices. available at http://site.ieee.org/pes-rrpasc/files/2019/04/National-Grid-Security-of-Supply-International-Review-Final-IEEE-v2.pdf.

NERC (2019). Generating Availability Data Systems: Data Reporting Instructions. Instruction manual.

NERC (2022). Generating Availability Data System (GADS) . https://www.nerc.com/pa/RAPA/gads/Pages/GeneratingAvailabilityDataSystem-(GADS).aspx. Accessed: 2022-03-14.

Newbery, D., Gissey, G. C., Guo, B., and Dodds, P. E. (2019). The Private and Social Value of British Electrical Interconnectors. *Energy Policy*, 133:110896.

NGESO (2017). Future energy scenarios. Technical report, National Grid. available at https://www.nationalgrid.com/document/138976/download.

NGESO (2019). Winter Outlook 2019/20. Technical report, National Grid. Available at https://www.nationalgrideso.com/document/154166/download.

NGESO (2020a). Decarbonizing heating for a net zero energy system. https://www.nationalgrideso.com/news/decarbonizing-heating-net-zero-energy-system. Accessed: 2021-12-14.

NGESO (2020b). Future Energy Scenarios 2020. Technical report, NGESO. Available at https://www.nationalgrideso.com/document/173821/download.

NGESO (2021a). Future Energy Scenarios 2021. Technical report, NGESO.

NGESO (2021b). Future Energy Scenarios 2021 data workbook. https://www.nationalgrideso.com/document/199971/download. Accessed on: 01-03-22.

NGESO (2021c). National Grid ESO Electricity Capacity Report. Technical report, NGESO. Available at https://www.emrdeliverybody.com/Capacity%20Markets%20Document%20Library/Electricity%20Capacity%20Report%202021.pdf.

NGESO (2022). Historic Demand Data. https://demandforecast.nationalgrid.com/efs_demand_forecast/faces/DataExplorer. Accessed on: 01-03-22.

Ofgem (2013). Electricity Capacity Assessment Report 2013. Technical report, Ofice of Gas and Electricity Markets. Available at https://www.ofgem.gov.uk/sites/default/files/docs/2013/06/electricity-capacity-assessment-report-2013_0.pdf.

Ofgem (2018). CONSUMER ENGAGEMENT IN THE ENERGY MARKET 2018. Technical report, Ofgem. Available at https://www.ofgem.gov.uk/sites/default/files/docs/2018/10/consumer_engagement_survey_2018_report_0.pdf.

Ofgem and DECC (2013). The Value of Lost Load (VoLL) for Electricity in Great Britain. Technical report, Ofgem. Available at https://www.ofgem.gov.uk/sites/default/files/docs/2013/07/london-economics-value-of-lost-load-for-electricity-in-gb_0.pdf.

Pang, C. and Wood, A. (1975). Multi-area generation system reliability calculations. *IEEE Transactions on Power Apparatus and Systems*, 94(2):508–517.

Pfenninger, S. and Staffell, I. (2015). www.renewables.ninja. https://www.renewables.ninja/. Accessed on: 01-03-22.

Richards, A. and Ong, K.-A. (2019). Average Cold Spell (ACS) Methodology. Technical report, NGESO. Available at https://www.emrdeliverybody.com/Lists/Latest%20News/Attachments/189/S\C4L12%20ACS%20Methodology.pdf.

Rienecker, M., Suarez, M., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M., Schubert, S., Takacs, L., Kim, G.-K., Bloom, S., Chen, J.,

Collins, D., Conaty, A., Da Silva, A., Gu, W., Joiner, J., Koster, R., Lucchesi, R., and Woollen, J. (2011). MERRA: NASA's Modern-Era Retrospective analysis for Research and Applications. *Journal of Climate*, 24:3624–3648.

Ringlee, R. J. and Wood, A. J. (1969). Frequency and Duration Methods for Power System Reliability Calculations: II - Demand Model and Capacity Reserve Model. *IEEE Transactions on Power Apparatus and Systems*, PAS-88(4):375–388.

Rootzén, H. and Tajvidi, N. (2006). Multivariate generalized pareto distributions. *Bernoulli*, 12(5):917–930.

Rumpf, J. and Bjørnebye, H. (2019). Just How Much Is Enough? EU Regulation of Capacity and Reliability Margins on Electricity Interconnectors. *Journal of Energy & Natural Resources Law*, 37(1):67–91.

Sanchez, N. (2022). `riskmodels`: a library for univariate and bivariate extreme value analysis (and applications to energy procurement). https://pypi.org/project/riskmodels/2.2.0/.

Sanchez, N., Dent, C. J., and Wilson, A. L. (2020). Quantifying the reliability contribution of interconnectors in the britain - ireland power system using a hindcast approach. In *2020 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, pages 1–6.

Sarykalin, S., Serraino, G., and Uryasev, S. (2008). Value- at-Risk vs Conditional Value-at-Risk in Risk Management and Optimization. *Tutorials in Operations Research*.

Savage, L. J. (1951). The Theory of Statistical Decision. *Journal of the American Statistical Association*, 46:56–67.

Schröder, T. and Kuckshinrichs, W. (2015). Value of Lost Load: An Efficient Economic Indicator for Power Supply Security? A Literature Review. *Frontiers in Energy Research*, 3:55.

Seelye, H. (1949). A Convenient Method for Determining Generator Reserve. *Transactions of the American Institute of Electrical Engineers*, 68:1317–1320.

Sensfuß, F., Ragwitz, M., and Genoese, M. (2008). The merit-order effect: A detailed analysis of the price effect of renewable electricity generation on spot market prices in Germany. *Energy Policy*, 36(8):3086–3094.

Sharifnia, E. and Tindemans, S. (2022). Multilevel Monte Carlo with Surrogate Models for Resource Adequacy Assessment. In *International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*.

Sheehy, S., Edwards, G., Dent, C. J., Kazemtabrizi, B., Troffaes, M., and Tindemans, S. (2016). Impact of High Wind Penetration on Variability of Unserved Energy in Power System Adequacy. In *2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, pages 1–6.

Shipley, R. B., Patton, A. D., and Denison, J. S. (1972). Power Reliability Cost vs Worth. *IEEE Transactions on Power Apparatus and Systems*, PAS-91(5):2204–2212.

Sibuya, M. (1960). Bivariate Extreme Statistics, I. *Annals of the Institute of Statistical Mathematics*, 11(3):195–210.

Singh, C. and Deng, Z. (1991). A new algorithm for multi-area reliability evaluation—Simultaneous decomposition-simulation approach. *Electric Power Systems Research*, 21(2):129–136.

Singh, C., Jirutitijaroen, P., and Mitra, J. (2018). *Electric Power Grid Reliability Evaluation: Models and Methods*. John Wiley & Sons, Ltd.

Singh, C. and Mitra, J. (1997). Composite system reliability evaluation using state space pruning. *IEEE Transactions on Power Systems*, 12(1):471–479.

Sioshansi, R., Denholm, P., Arteaga, J., Awara, S., Bhattacharjee, S., Botterud, A., Cole, W., Cortés, A., Queiroz, A. d., DeCarolis, J., Ding, Z., DiOrio, N., Dvorkin, Y., Helman, U., Johnson, J. X., Konstantelos, I., Mai, T., Pandžić, H., Sodano, D., Stephen, G., Svoboda, A., Zareipour, H., and Zhang, Z. (2022). Energy-storage modeling: State-of-the-art and future research directions. *IEEE Transactions on Power Systems*, 37(2):860–875.

Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publications del'Institut de Statistique de l'Université de Paris*, 8:229–231.

Smith, J. Q. (2010). *Utilities and rewards*, page 62–102. Cambridge University Press.

Smith, S. (1934). Spare capacity fixed by probabilities of outage. *Electrical world*, 103:222–225.

Staffell, I. and Pfenninger, S. (2016). Using Bias-Corrected Reanalysis to Simulate Current and Future Wind Power Output. *Energy*, 114:1224 – 1239.

Staffell, I. and Pfenninger, S. (2018). The increasing impact of weather on electricity supply and demand. *Energy*, 145:65 – 78.

Steiner, M. and Meißner, M. (2018). A user's guide to the galaxy of conjoint analysis and compositional preference measurement. *Marketing ZFP*, 40:3–25.

Stenclik, D., Cole, W., Figueroa, A., Stephen, G., Tuohy, A., and A., B. (2021). Redefining Resource Adequacy for Modern Power Systems. Technical report, ESIG. Available at https://www.esig.energy/resource-adequacy-for-modern-power-systems.

Stephen, G., Tindemans, S. H., Fazio, J., Dent, C. J., Figueroa Acevedo, A., Bagen, B., Crawford, A., Klaube, A., Logan, D., and Burke, D. (2022). Clarifying the Interpretation and Use of the LOLE Resource Adequacy Metric. In *International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*.

Stephenson, A. (2003). Simulating multivariate extreme value distributions of logistic type. *Extremes*, 6:49–59.

Stickler, A. and Brönnimann, S. (2011). Significant bias of the NCEP/NCAR and twentieth-century reanalyses relative to pilot balloon observations over the West African Monsoon region (1940-1957). *Quarterly Journal of the Royal Meteorological Society*, 137:1400 – 1416.

Telson, M. L. (1975). The economics of alternative levels of reliability for electric power generation systems. *The Bell Journal of Economics*, pages 679–694.

Thornton, H. E., Scaife, A. A., Hoskins, B. J., and Brayshaw, D. J. (2017). The relationship between wind power, electricity demand and winter weather patterns in great britain. *Environmental Research Letters*, 12(6):064017.

Tindemans, S. and Strbac, G. (2020). Accelerating system adequacy assessment using the multilevel Monte Carlo approach. *Electric Power Systems Research*, 189:106740.

Tindemans, S. H. and Strbac, G. (2015). Visualising risk in generating capacity adequacy studies using clustering and prototypes. In *2015 IEEE Power Energy Society General Meeting*, pages 1–5.

Tindemans, S. H., Woolf, M., and Strbac, G. (2019). Capacity Value of Interconnection Between Two Systems. In *IEEE Power and Energy Society General Meeting (PES GM)*.

UK Government (2022). Energy Trends: UK renewables . https://www.gov.uk/government/statistics/energy-trends-section-6-renewables. Accessed on: 01-03-22.

Vassell, G. S. and Tibberts, N. (1972). Analysis of generating-capacity reserve requirements for interconnected power systems. *IEEE Trans. Power Apparatus and Systems*, PAS-91(2):638–649.

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., and Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology*, 60(3):158–189.

Wheatcroft, E. and Dent, C. and Wilson, A. (2022). Rescaling of Historic Electricity Demand Series for Forward-Looking Risk Calculations. In *International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*.

Wilson, A. L. and Zachary, S. (2019). Using extreme value theory for the estimation of risk metrics for capacity adequacy assessment. *arXiv preprint arXiv:1907.13050*.

Wilson, A. L., Zachary, S., and Dent, C. J. (2018). Use of Meteorological Data for Improved Estimation of Risk in Capacity Adequacy Studies. In *2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, pages 1–6.

Zachary, S. and Dent, C. (2014). Estimation of Joint Distribution of Demand and Available Renewables for Generation Adequacy Assessment.

Zachary, S. and Dent, C. J. (2012). Probability theory of capacity value of additional generation. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 226(1):33–43.

Zachary, S., Wilson, A., and Dent, C. (2021). The integration of variable generation and storage into electricity capacity markets. *Energy Journal*.