Department for
Energy Security
& Net Zero

Exploring Reliability Standard Metrics in a Net Zero Transition

DESNZ research paper number: 2023/050

November 2023

Any enquiries regarding this publication should be sent to us at: remamailbox@energysecurity.gov.uk

Contents

Executive Summary

Capacity adequacy is one of the key areas for reform being assessed in the Review of Electricity Markets Arrangements (REMA) Programme with the reliability standard critical to this assessment.

The Great Britain electricity system is set to undergo significant change in the coming years as the system evolves to achieve decarbonisation targets whilst meeting significant increases in demand. Ensuring capacity adequacy in this future electricity system will be a different prospect to ensuring capacity adequacy today as demand changes in volume and nature, and intermittent renewables start to dominate our electricity generation.

The reliability standard is a key component of capacity adequacy. It sets the level of reliability that the electricity system must meet and is used within the Capacity Market to define how much capacity needs to be procured within each auction. Therefore, it is critical to understand whether the current reliability standard metric for use in the Capacity Market will be appropriate in the future. Considering the prevailing and anticipated generation landscape, the current reliability standard, 3 hours Loss of Load Expectation (LOLE) per year, may not be sufficient to reflect these new risks as we transition to a fully decarbonised power system.

LCP Delta and Frontier Economics were commissioned by the Department for Energy Security and Net Zero (DESNZ) to assess whether the current reliability standard metric remains an appropriate metric for capacity adequacy as GB transitions to a fully decarbonised electricity system

Frontier Economics (Frontier) and LCP Delta (LCP) were engaged by DESNZ to examine possible changes to the Reliability Standard for Great Britain (GB). The project examined if the current reliability standard metric remains an appropriate measure for capacity adequacy risk as we move towards a net zero electricity system and explored whether alterative metrics are more appropriate. To carry out this assessment, the study was split into two stages:

- In the first stage of this study, Frontier carried out a review of selected literature about capacity adequacy and associated metrics. This enables DESNZ to gain an understanding of how future stress events may change from today, the challenges and risks in using LOLE as the adequacy metric, and what other reliability standard metrics can be used to either replace LOLE or be used alongside it. From this research, some key hypotheses have been formed on how the reliability standard may need to change in the future to meet the new challenges that the GB electricity system's transition presents.
- In the second stage of the study, LCP has undertaken empirical analysis to test the findings from the literature review. This has been done by conducting modelling to assess different reliability standard options. The modelling first assesses the characteristics of system stress events, and how these change through to 2035 under two market scenarios (a DESNZ scenario and a NG ESO scenario). It then tests the implications of using alternative reliability standard metrics, including the level of capacity that needs to be procured.

The literature review finds that stress events are expected to change in the future and become more complex. As a result, there could be a case for considering a combination of metrics in the Reliability Standard.

A number of studies were reviewed to understand how stress events might change in the future and whether there is likely to be a need to consider alternative reliability metrics that could either be used in place of LOLE or be used alongside it.

The literature on reliability metrics is broad, with a number of different metrics used in different countries (e.g. Australia uses expected energy unserved, Mexico uses loss of load probability), and a LOLE target commonly applied (e.g. in many European countries and parts of the US). There is a clear recognition in the literature that reliability metrics may need to evolve, however, there is limited guidance as to how precisely that should be done. The following key findings were identified in the literature review:

- The nature of system stress events is expected to become more complex. As the system evolves, with a greater reliance on weather dependent renewables in particular, the system may experience a number of changes including larger reductions in supply than historically observed due to correlated renewables output; and longer periods of reduced supply due to risks of extended periods of low wind i.e. “dunkelflaute”.
- Increasingly, targeting a single LOLE metric may no longer be appropriate. As system stress events change, it could be that although a LOLE target can be met, other dimension of risks which are not captured by LOLE could increase or become more extreme. This could include an increase in the volume of energy unserved or an increase in the duration of stress events. The risks of extreme events could also increase, as LOLE only captures the expected (average) number of hours of stress events over the year. As a result, the loss of load costs faced by end-use consumers and system operators could be quite different for a given value of LOLE.
- A single superior metric was not identified from the literature. However, there are clear indications that a combination of metrics may be beneficial. There are some limited international examples, e.g. Belgium, which in addition to a LOLE measure also procures capacity to ensure that very rare but more extreme events are limited.
- While there appears to be a reasonably clear argument for considering multiple metrics, the literature does not provide a clear analytical framework for setting the target level of multiple metrics that strikes an optimal balance between the value of lost load and the cost of incremental additions to the capacity mix. Further development of the theoretical framework is therefore required.
- Ultimately, the extent to which these findings are applicable to GB will be a function of how our system is expected to evolve as part of the energy transition. Therefore, empirical analysis (including that carried out as part of this study) could help in determining whether there is a need for adopting multiple metrics and deciding on the optimal choice of metrics.

The modelling finds limited evidence to move away from LOLE as one of the reliability standard metrics at this time. However it does find that how key reliability metrics change in the future system depends on the scenario assumed

For the first part of the modelling, an assessment was made of how each of the key metrics identified through the literature review change over time. This was assessed over the period from 2025 to 2034, under both DESNZ's 2022 Net Zero Higher Demand (NZH) scenario and NG ESO's FES 2022 System Transformation (ST) scenario. In order to assess how other reliability metrics evolve over time on a consistent basis, the average total length of system stress events (LOLE) was fixed to 3 hours/year in line with the existing GB reliability standard. . The metrics modelled were:

Table 1: *Reliability Standard Metrics used in modelling.*

Metric	Definition
Loss of Load Expectation (LOLE)	The average number of hours per year in which it is expected statistically that supply will not meet demand and the system operator must take exceptional measures outside the normal operation of the system, including voltage reduction and partial rolling outages. Not all exceptional measures would have a significant noticeable effect on consumers. .
Loss of Load Frequency (LOLF)	The average number of events in a year where there is insufficient supply to meet demand.
Loss of Load Duration (LOLD)	The average length of an event in a year where there is insufficient supply to meet demand.
Expected Energy Unserved (EEU)	The amount of electricity demand – measured in MWh – that is expected not to be met by supply in a given year.
Normalised EEU	EEU normalised based on peak demand.
Critically Tight Events	The number of hours in a year for which it is expected that demand is over 99% of the total available supply.

In the two scenarios modelled, most metrics do not show significant change from 2025 to 2034. However, more change is observed across the metrics in the FES ST scenario compared to the DESNZ NZH scenario. For example, the Loss of Load Duration (LOLD) metric shows little change in the DESNZ NZH scenario but increases from around 3 hours/event in 2025 to over 5 hours/event in 2034 in the FES ST scenario. This is due to the nature of system stress events changing more in the FES ST scenario due to differences in the makeup of the capacity mix and the nature of demand in this scenario.

This highlights that what is assumed about the future of the energy system and how it evolves to reach a fully decarbonised system is highly important in determining whether there is a case for changing the reliability standard. A system with flatter daily demand due to more Demand Side Response (DSR) and a lower electricity for heating demand, such as the FES ST scenario, would more likely see benefits in changing reliability standard compared to a scenario that sees its stress events determined by higher daily peak demand and with large amounts of small peaking capacity on the system (such as the DESNZ scenario).

This stage of the modelling finds that it would be appropriate to keep LOLE as part of a combination of metrics making up the reliability standard. With the exception of LOLD (under certain scenarios) and EEU (including the normalised EEU), the modelling did not observe significant changes in most of the dimension of risks that alternative metrics aim to capture, when the current reliability standard of 3 hours LOLE is met. This indicates that many of the metrics modelled are not good candidates for an alternative reliability standard. However this should be kept under review as the direction of decarbonisation within the GB electricity system becomes more apparent.

A combined reliability standard metric based on Loss of Load Expectation (LOLE) and Loss of Load Duration (LOLD) could provide additional capacity adequacy in certain scenarios.

One of the key findings from the literature review is that a combination of metrics could be beneficial as the nature of system stress events become more complex. To test this finding, we have assessed the impact of a combined reliability standard based on LOLE and LOLD, both with a target of 3 hours (per year and per event respectively) based on maintaining 2025 levels. LOLD was chosen as the second metric here as it was the metric that showed the most significant change over time. This was only modelled for the FES scenario given LOLD only showed significant change in this scenario. Other metrics, such as volatility metrics identified in the literature or the normalised Expected Energy Unserved (EEU), could also warrant further investigation as a second metric but has not been modelled. Due to the nature of the modelling, using LOLD as a second, combined metric allowed insight into the impact of procurement on these other changing metrics as well. However, it is possible that additional information could be gathered on the exact relationship between the different metrics if they are modelled directly as second, combined metrics. If a combined reliability standard is taken forward as an option by DESNZ, a full review of the target level for that second metric would be needed but that is out of scope for this study.

The analysis shows that a combined reliability standard of LOLE and LOLD would provide additional capacity adequacy as it would result in more capacity being procured in the Capacity Market. From 2025 to 2034, this results in the procurement of an additional 13% of derated capacity (additional 7GW in 2034 the FES ST scenario) compared to retaining just the existing reliability standard. This additional capacity results in LOLD meeting the 3 hour/event target but leads to LOLE decreasing over time. This is because as the average duration of an event (LOLD) stays at the target level of around 3 hours/event while the frequency of an event (LOLF) decreases. This shows that targeting both LOLE and LOLD will decrease the frequency of events in addition to decreasing the length of events.

However, procuring this additional capacity comes at an additional cost in the Capacity Market. As an illustration based on the net CONE (cost of new entry) assumption currently in use in the Capacity Market, procuring this level of additional capacity would cost an additional £343m in 2034 on top of the costs of maintaining the existing reliability standard. It should be noted this cost reflects the additional capital cost of procuring the additional capacity within the Capacity Market only and does not take into account the full second order impact on operational system costs. The reduction in risk as a result of procuring additional capacity could affect system costs in different ways which may bring net benefits to the whole system.

In addition to changing the amount of capacity that needs to be procured, changing the reliability standard could also affect the type of capacity mix procured by the Capacity Market. This is because it affects Equivalent Firm Capacities (EFCs), which are used to calculate derating factors for intermittent generation and for limited duration storage. The analysis shows that a combined LOLE and LOLD reliability standard, allows storage to make a greater contribution to capacity adequacy and leads to higher storage derating factors. Shorter system stress events as a result of a lower LOLD mean that storage plants can contribute to negating them more frequently. This means that the EFCs for storage plants increase over time and are higher than in the baseline scenario. A higher storage derating factor could lead to more storage being procured through the Capacity Market, other things being equal.

Overall, the modelling results support the finding from the literature review that a combined reliability standard could provide additional capacity adequacy to mitigate risks. Further empirical analysis is needed to fully determine what a future combined reliability standard should look like given some of the elements needed to determine this are out of scope of this study. Further work could include modelling of alternative scenarios, a more detailed evaluation of system cost impacts and/or a refresh of analysis previously completed to economically determine an updated LOLE target or new combined target metric.

1. Introduction and Scope

1.1. Background and Context

The Government published its Review of Electricity Market Arrangements (REMA) consultation in July 2022[1] outlining a range of options for reforms to electricity markets. The aim of the REMA programme is to ensure that the market is fit for the purpose for the future by delivering reform to electricity market arrangements that facilitate the full decarbonisation of the electricity system by 2035, subject to security of supply, whilst being cost effective for consumers.

One of the areas considered for reform within the REMA programme is Capacity Adequacy – identifying the most appropriate capacity adequacy mechanism and wider adequacy framework to address future challenges. Ensuring capacity adequacy in our future electricity system will be a different prospect to ensuring capacity adequacy today. This is mainly due to two factors; increasing demand due to electrification in other sectors such as transport and heat and increasing penetration of intermittent. As a result, the REMA programme includes options for changes to the system to continue to ensure capacity adequacy in the future. This includes options for reforming or replacing the existing capacity mechanism to better support firm low carbon technologies.

As part of this assessment, it is important to understand whether the current reliability standard metric for use in the Capacity Market will be appropriate in the future. Considering the prevailing and anticipated generation landscape, loss of load expectation (LOLE) may not be sufficient on its own to reflect the risks as we transit towards a fully decarbonised power system.

Given the possible future changes to the system, Frontier Economics (Frontier) and LCP Delta (LCP) were engaged by the Department for Energy Security and Net Zero (DESNZ) to examine the current Reliability Standard metric for Great Britain (GB). DESNZ asked Frontier and LCP to examine whether this metric remains an appropriate metric for capacity adequacy as GB transitions to a net zero electricity system with a flexible, low carbon capacity mix.

1.2. The Capacity Market and Reliability Standards

Since its introduction through the Electricity Market Reform (EMR) programme in 2014, the Capacity Market (CM) has ensured that Great Britain maintains and brings forward sufficient capacity to deliver a secure electricity supply. It provides payments to capacity providers, in return for which they must be available to produce energy when the system is tight to avoid system stress events (when demand is greater than supply). The Capacity Market works

---

[1] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1098100/review-electricity-market-arrangements.pdf

through a competitive auction process where participants bid to secure a contract 1-year (known as T-1 auction) and 4-yerars ahead (T-4 auction).

The reliability standard is a key component of the Capacity Market. It sets the level of reliability that the electricity system must meet and is used within Capacity Market auctions to define how much capacity needs to be procured within each auction. The existing reliability standard was designed in 2013 to inform targets for the Capacity Market in an electricity system still heavily reliant on flexible and dispatchable fossil-fuel power generation. Currently, the Government sets a Reliability Standard of 3 hours Loss of Load Expectation (LOLE) per year.[2] The standard is determined as the ratio of the Value of Lost Load (VOLL) and the Cost of New Entry (CONE).[3]

1.3. Project Overview

For this project, DESNZ had commissioned Frontier/LCP to gather evidence in helping to examine if LOLE will remain an appropriate reliability standard metric for our future electricity system and to explore whether alternative metrics are more appropriate. The project aims to:

- Examine the appropriateness of the current reliability standard metric in reflecting future risks as we transition towards a net zero electricity system.
- Explore the range of alternative metrics that could either replace or be used alongside LOLE, and critically assess their benefits and limitations.
- Consider the implication from using an alternative metric on the process of target setting and in determining the level of capacity procurement.
- Provide quantitative insights on the impacts of using alternative reliability standard metrics on overall security of supply.

To achieve this, qualitative and quantitative analysis has been undertaken to gain insight into whether the reliability standard metric should be changed in the future. This includes:

- An examination of how stress events will change in the future and to the extent this impacts the appropriateness of LOLE as a metric
- An exploration of the challenges and risks for using LOLE as a metric
- An examinations of other reliability standard metrics
- Analysis on the impact on security of supply of using LOLE and alternative metrics.

The outputs of this study will help DESNZ in developing thinking in wider capacity adequacy reforms and in identifying further areas of research.

It should be noted that this study does not consider the ability for system operators to manage different dimensions of risks. This would also be a factor in choosing a reliability standard

---

[2] Annex_C_-_reliability_standard_methodology.pdf (publishing.service.gov.uk)

[3] This is based on the cost of a new OCGT plant - emr_consultation_annex_c.pdf (publishing.service.gov.uk)

metric, for example it may be easier to address longer but shallower events (for example through voltage reduction or rolling outages) compared to short but deep events. This is out of scope of this study as it is more of a practical consideration for the system operator to consider alongside DESNZ when choosing a new metric as opposed to something to test in identifying possible alternative metrics.

1.4. Project Approach

In delivering this project Frontier and LCP have split the project into two key stages; a literature review and a modelling exercise. In the first stage of this study, Frontier has carried out a review of selected literature to better understand how the transition towards a net zero electricity system could impact the nature of future system stress events and the reliability metrics used for ensuring capacity adequacy. In particular, Frontier has reviewed the literature with the objective to answer the following questions:

- How may future system stress events differ from those experienced historically and today?
- What are the challenges and risks in using LOLE as the single adequacy metric for setting the reliability standard?
- What other reliability standard metrics can be used to either replace LOLE or be used alongside LOLE? Which of these metrics are most robust to different scenarios in capturing risks?

To consider these questions, a number of studies and reports have been reviewed reflecting the wide range literature on this topic. These studies have then been used to identify potential alternative metrics and approaches to test in the modelling stage of the project.

In the second stage of the study, LCP's Unserved Energy Model (UEM) has been used to assess system stress event characteristics and the implications on the use of different reliability standard metrics.

The modelling first assesses the characteristics of system stress events, and how they change over the next 10 years out to 2035 in a transitioning system for two different scenarios. We have modelled it up to 2035 as the government has a target for the power sector to be fully decarbonised by 2035, subject to security of supply. We assume the system is expected to change relatively little beyond 2035 in terms of its security of supply behaviour.

The modelling shows how the potential metrics will change over time as renewable penetration and demand increase. The modelling then tests the implications of using different reliability standard metrics to understand the impact alternative metrics could have on the capacity that needs to be procured within the capacity market. The impacts that these metrics could have on the deratings of different technologies is also assessed.

2. Literature Review

In the section we summarise the findings of our literature review. These findings have also informed the quantitative modelling undertaken in the second phase of this study. Table 1 below lists the studies that we have reviewed. We refer to these studies throughout this section.

It is worth noting that while there are numerous studies discussing reliability standard metrics, there are relatively few studies tackling how reliability standards *may need to evolve* in the face of decarbonisation and climate change. Some key reports which we have reviewed that take this forward-looking perspective specifically are the following:

- The National Grid ESO December 2022 report on resource adequacy in the 2030s (ESO 2022);
- The Energy System Integration Group (ESIG)'s report on redefining resource adequacy for modern power systems (ESIG 2021). ESIG is a US based non-profit educational organisation whose mission is to chart the future of grid transformation and energy systems integration;
- The Public Utility Commission of Texas's 2023 consultation on reliability standards for the ERCOT market (PUC 2023) with a particularly useful response from London Economics International (London Economics International 2023). ERCOT is the Electricity Reliability Council of Texas, a US organisation that operates Texas' electricity grid. Texas currently does not have a reliability standard in place; the consultation is being held after Winter Storm Uri in 2021 led to widespread blackouts[4]. The process is likely to conclude in 2024 (PUC, 2023a).

The remainder of this section is structured as follows:

- First, we discuss the evidence on how the nature of future system stress events is likely to change;
- Second, given this potential change in the nature of system stress events, we discuss the implications for the current approach of setting a reliability standard based on LOLE;
- Third, we consider a range of alternative metrics to setting a reliability standard; and
- Finally, we set out the implications of this review, including for the modelling undertaken in the second phase of this study, and some suggested areas where further conceptual thinking is required.

---

[4] Utility Dive, March 2023," Texas regulators question common reliability metric as they pursue new standard for ERCOT grid", accessed August 2023. https://www.utilitydive.com/news/texas-ercot-reliability-standard-puct/646129/#:~:text=Texas%20does%20not%20have%20a,250%20people%20in%20the%20state.

Table 2: *List of studies reviewed*

Short form citation	Study	Source
ACER (2022)	Security of EU electricity supply in 2021: Report on Member States approaches to assess and ensure adequacy	source
BMWK (2019)	Definition and monitoring of security of supply on the European electricity markets	source
Beyza (2021)	Comparative Evaluation of the Reliability and Vulnerability of Electrical Networks with a High Share of Renewable Generation	source
CERRE (2023)	Building Resilience in Europe's Energy System	source
DECC (2013)	Annex C: Reliability Standard Methodology	source
ERCOT (2014)	Estimating the Economically Optimal Reserve Margin in ERCOT	source
ESIG (2021)	Redefining Resource Adequacy for Modern Power Systems	source
ESO (2022)	Resource adequacy in the 2030s	source
Garip (2022)	Power system reliability assessment – A review on analysis and evaluation methods	source
Heylen et al. (2018)	Review and Classification of Reliability Indicators for Power Systems with a High Share of Renewable Energy Sources	source
London Economics (2013)	The Value of Lost Load (VoLL) for Electricity in Great Britain–	source
London Economics International (2023)	Economic considerations for setting reliability standards for the wholesale power market in Texas	source
NERC (2016)	2016 Probabilistic Assessment	source
PUC (2023)	Project No. 54584 – Reliability Standard for the ERCOT Market – Open Meeting Cover Sheet Memorandum	source filings
PUC (2023a)	Filing of August 21st 2023	source
Vistra Corp (2023)	Project No. 54584 – Vistra Corp.'s Reply Comments	source

2.1. The nature of future stress events is changing

In this section we set out our findings related to how the nature of future system stress events is likely to change. This section is structured as follows:

- First, we define four key dimensions that can be used to classify system stress events. These dimensions will be used in the following sections to describe how system stress events are changing and to identify reliability metrics that can be used to target each dimension.
- Second, we summarise the key drivers of reliability concerns in traditional (thermal based) electricity systems with production of power from large centralised power stations, and describe the nature of system stress events according to the four dimensions. Understanding historical drivers of reliability concerns helps in framing how these drivers may change in the future.
- Third, we summarise key changes that are likely to influence the future electricity system in Great Britain (GB), in particular those changes related to the decarbonisation of the electricity system. These changes will impact future system stress events.
- Finally, we explain the implications that these changes to the system may have on the drivers of reliability concerns in future and therefore the nature of future system stress events.

System stress events can be classified according to four key dimensions

According to the literature, system stress events can be characterised in a number of different dimensions. For example:

- Heylen et al. (2018) sets out an overview of *“the characteristics and the scope of [reliability] indicators”*, structuring these into a framework of **frequency**, **probability**, **magnitude**, and **duration**;
- ESIG (2021) also set out a framework to assess risks in terms of **size**, **frequency**, **duration**, and **timing**; and
- The Electric Reliability Council of Texas (ERCOT) has recommended, in a consultation on the reliability standard, *“that the standard be defined by a three-part framework that touches on multiple metrics, including establishing limits on the **duration**, **frequency**, and **magnitude** of loss of load events”* (PUC, 2023).

The precise manner in which the different dimensions are described varies slightly, but there is a high degree of consistency among the papers we reviewed. For ease, in the remainder of this discussion we refer to the dimensions as we have set out in the table below. These dimensions can be used to describe how system stress events are expected to change, as well as to identify reliability metrics that target those dimensions. In the following sections we will use this classification to describe how system stress events are expected to change in the future.

Table 3: *Dimensions of stress events*

Key characteristics of stress events	**Meaning**
Frequency of events	How often events are likely to occur (per year)
Duration	How long events are likely to be (in hours)
Size	How large events are likely to be (in MWh)
Volatility	How much the above measures vary across events that are expected to occur

Historic drivers of reliability concerns and traditional system stress events

Historically, the GB electricity system has been based around the production of power from large centralised carbon intensive power stations. Although this is rapidly changing, the historic characteristics of the system are core to how resource adequacy risks have been considered to date.

In such a system, the key drivers of reliability concerns were typically **unplanned outages of power station units or interconnectors over periods of peak demand**, which were considered the periods in which the system would be most vulnerable. **Fuel availability was generally not considered a problem** as supplies of gas, coal, and uranium were regarded as plentiful and unconstrained. As a result, periods when the system was stressed ("system stress events") were thought most likely to occur due to the coincidence of multiple power station outages with peak demand.

Traditional assessments of resource adequacy typically considered that **plant outages were independent (uncorrelated) events** and that mechanical failures of generating equipment occurred at random (ESIG, 2021). The implication of this was that the probability of multiple failures occurring simultaneously was low.

Because of this, traditional system stress events were expected to be relatively homogenous, of short duration, relatively rare, and small in size. System stress events were also generally considered to have low serial correlation, i.e. the likelihood of a system experiencing stress in a particular hour, is relatively independent of whether the system was experiencing stress in the previous hour.

A number of factors are likely to change the GB electricity system

The GB electricity system is expected to change in the future. Our literature review (including ESO, 2021) identified five key factors that will likely have an impact on the GB electricity system and therefore future system stress events, shown in Table 3. Below the table we describe in more detail how these changes are likely to impact the future electricity system. In the next section we summarise the implications of these changes on system stress events.

Table 4: *List of factors that are likely to affect the future GB electricity system*

Factors	Likely impact on the GB electricity system
Increasing demand for electricity	Higher load on the electricity system, and potentially larger peaks (in MW terms) Potentially higher level of 'firm capacity' required due to potential increase in system peak demand and higher value placed by customers on secure supplies due to new uses of electricity
Increasing penetration of intermittent renewables	Greater fluctuations in output, and increased risks of extended periods of low output
Increasing need for flexibility to accommodate intermittency	Increase in smaller units of dispatchable power, and increased reliance on storage e.g. batteries
Increasing interdependence with neighbouring countries	Lower security of supply risks for uncorrelated events Higher risks for coincident events
More extreme weather patterns due to climate change	In combination with electrification and increasing penetration of renewables, can make systems more vulnerable

Increasing demand for electricity

As outlined in the Net Zero Strategy,[5] electricity demand is expected to rise by 40-60% by 2035 due to electrification in sectors such as transport and heat. This has a number of implications for system security:

- Along with overall increases in electricity demand, system peak demand may increase, increasing the level of 'firm capacity' required;
- Customers may become more flexible in their electricity demand and responsive to price, and as a result flattening peaks in demand; and
- The value that customers place on secure supplies of electricity may change as electricity is used to meet essential needs such as heat and transport where the value associated with loss of load may be higher.

---

[5] Net Zero Strategy, October 2021, BEIS, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1033990/net-zero-strategy-beis.pdf p. 82

Increasing penetration of intermittent renewables

The UK government has set out a target to decarbonise its electricity system by 2035,[6] including ambitious targets to increase the volume of renewable technologies, in particular offshore wind. Therefore, the system will be operating with much greater volumes of generation that is weather-dependent. This means that:

- Output across wind or solar farms are likely to be correlated, increasing the potential for significant fluctuations in output;
- Output is also likely to be serially correlated, i.e. if output is low in one period then this increases the probability of low output in the next period, increasing the risks of extended periods of system stress.

Increasing need for flexibility

To accommodate the increasing penetration of intermittent renewables, the system will require more flexible capacity to be provided from other sources. This is likely to lead to an increasing role for storage (ESO, 2021) and in particular battery storage,[7] as well as other sources of dispatchable generation and flexible demand. Increasing decentralisation will see much of this flexibility as small units located on distribution networks.

Increasing interdependence with neighbouring countries

A significant expansion of new interconnection capacity is expected over the next decade, increasing the level of interdependence with neighbouring countries (ESO, 2021). This will help reduce security of supply risks when system stress events are uncorrelated; and increase exposure to coincident events, e.g. due to extreme weather patterns extending across Western Europe.

More extreme weather patterns due to climate change

Climate change drives more extreme weather patterns. If global warming is not limited, the IPCC expects the frequency of extreme temperature events to increase by 50% relative to the current level, and heavy precipitation over land to increase by 15% (CERRE, 2023). Electricity systems more dependent on weather-dependent renewable sources will be increasingly vulnerable to such extreme events (CERRE, 2023).

System stress events are expected to become more complex

The changes to the GB electricity system summarised in the previous section will lead to changes to the nature of system stress events. Overall, system stress events are expected to become more complex as more of the dimensions we have set out become more important. This conclusion is supported by a range of studies, for example ESO (2021) and ESIG (2021).

---

[6] UK Government Press Release, Plans unveiled to decarbonise UK power system by 2035 https://www.gov.uk/government/news/plans-unveiled-to-decarbonise-uk-power-system-by-2035

[7] Solar Quarter, UK Battery Storage Market Poised for Exponential Growth, Set to Reach 24 GW by 2030, 24.04.2023, accessed 14.07.2023 https://solarquarter.com/2023/04/24/uk-battery-storage-market-poised-for-exponential-growth-set-to-reach-24-gw-by-2030/

The key implications of the changes described above for system stress events are the following:

- On the one hand, the **impact of plant outages (as share of total generating capacity) may reduce.** Much like large thermal power stations, wind and solar plants are also expected to have unplanned outages due to mechanical or electrical failures. However, the impact of these random *uncorrelated* plant outages on the system is likely to reduce given the small and modular nature of renewable plants, i.e. unplanned outages are still likely to occur but their impact as a share of total generating capacity is likely to be much lower.
- On the other hand, **fuel availability is likely to become a much more significant issue**. There are three reasons for this:
    - Given the increase in weather dependent resources, changes in output across the generating capacity due to changes in the weather are likely to be more correlated than historically. In other words, in future **the system may face higher risks of large reductions in output due to the weather** than would have been expected due to multiple power station outages;
    - Given the weather in one hour is linked to the weather in the following hour, if wind or solar output is low in one period then this increases the probability of low output in the next period, **increasing the risks of extended periods of system stress** (i.e. changes in renewables output exhibit serial correlation).
    - An increasing dependence on storage capacity to respond to changes in the output from variable renewables means that **the ability of the system to be able to respond to sudden changes in output also depends on what has happened in the previous periods**, i.e. the state of charge of storage capacity at a point in time will be critical (ESIG, 2021).

  In other words, in future the system may not only be constrained in the amount of generating capacity that is available, but also by the energy that is available.
- There is likely to be **an increasing focus on off-peak as well as peak periods**, since large reductions in correlated renewable output can lead to an imbalance between supply and demand in periods of relatively low demand. For example, in California in August 2022, *"involuntary rolling outages … occurred late in the evening after the sun had set and solar resources dropped off, several hours after peak load occurred in the middle of the day."* (ESIG, 2021)

As a consequence of these changes, some of the dimensions of system stress events are likely to become more important. The table below summarises how each of these dimensions could change when considered *in isolation* given the drivers of changes that we have identified. However, it is important to note that once a reliability standard is imposed on the system, some of these dimensions will be constrained and therefore will not *all* move in the direction indicated in the table. For example, as explained in Section 2.3,[8] LOLE can be thought of as the average duration of a stress event multiplied by the number of stress events

---

[8] See explanation and formula in the sub-section titled 'LOLF (loss of load frequency)'.

in a year. If a LOLE standards is imposed on the system, if the average duration increases the number of stress events is constrained to decrease (and vice versa). This is ultimately an empirical question which depends on the assumptions made around the factors that drive stress events and the reliability standards imposed (for example, a study by AFRY for ESO (2022) found that given a LOLE standard stress events will become less frequent but longer in duration (and larger in size).

We elaborate on the impact that these changes are expected to have on each of the four dimensions in turn.

Table 5: *Expected changes to system stress events by dimension (when considered in isolation)*

Dimension	Past system stress events	Likely future system stress events
Frequency of events	Less frequent (as events less correlated)	More frequent (as events more correlated)
Duration	Shorter duration	Longer duration
Size	Smaller	Larger
Volatility	Relatively homogenous events	More volatile events

Frequency of events

Previously, system stress events were driven by random, uncorrelated outages during peak load hours. Therefore, it was quite rare for enough outages of individual power station units to occur simultaneously such that a risk of loss of load was created for the system as a whole.

However, in the future the nature of the supply mix (in particular its correlated output) may mean that supply fluctuations will be more frequent, leading to a higher frequency of system stress events.

Duration of events

Previously, system stress events have been driven by uncorrelated generator outages combined with peak periods, and hence they were expected to be relatively short in duration. Demand typically has had relatively predictable and short peaks.

In the future, given the fact that changes in weather, and hence output of wind and solar, are likely to be serially correlated, reductions in renewables output that can increase the risk of loss of load may last for an extended period of time.

Unfavourable weather patterns will be an important driver of future system stress events. In particular, ESO 2021 considers that weather patterns extending across North-West Europe and resulting in prolonged periods of low wind will be *"the most challenging situations"*.

> *"[W]eather patterns will be the dominant driver of stress periods in a fully decarbonised power system … weather patterns extending across North-West Europe that result in prolonged periods of low wind during winter ... Could lead to much longer periods of*

> *system tightness. […] Increased interconnector capacity means the future system becomes much more exposed to NW Europe-wide periods of low generation"* (ESO, 2022).

GB is likely to face these challenging situations. Periods of low generation due to unfavourable weather patterns (***Dunkelflaute***) have been found to occur at least 5 times per year over 1985-2016 - we provide more details in the box below. In future, as the generation mix evolves, the system is expected to become more exposed to these types of events.

The change in the expected duration of system stress events is likely to be material. ESO's report on Resource Adequacy in the 2030s shows a significant increase in *critically tight periods*[9] between 2025 and 2040*:* in 2025 the longest critical tight period is expected to be between 8-15 hours; by 2040 the longest critically tight period is expected to increase to 51-75 hours (in the Consumer Transformation scenario).

Dunkelflaute in GB over 1985-2016

Dunkelflaute refers to a period of winter weather with low light and little to no wind. This *is "a fairly common phenomenon in Northern Europe, occurring between 50 and 100 hours each November, December and January"*[10]. It tends to coincide with times when energy use is likely to be high due to cold temperatures.

The figure below shows, based on simulations, how often Dunkelflaute occurred in the UK in the period 1985-2016.[11] The chart shows that periods of low generation (less than 40% of capacity for both solar and wind), lasting up to 4 days, occurred 5 – 10 times per year over this period (represented by the yellow area in the chart).

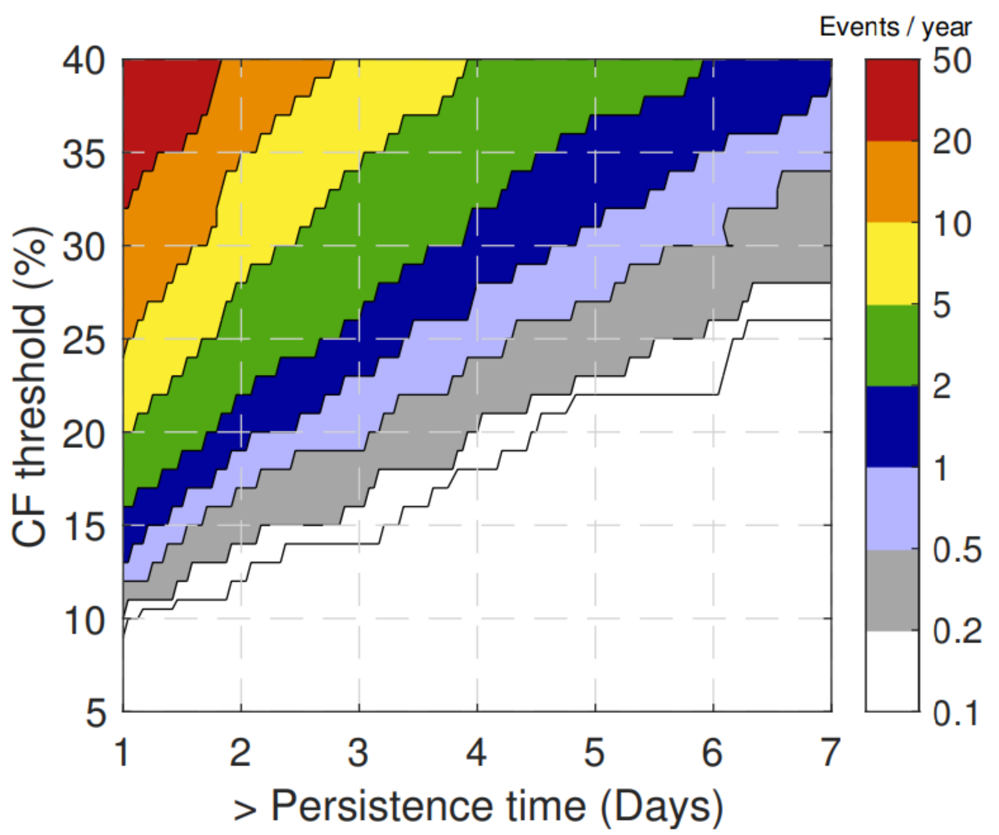
---

[9] A critically tight period is an hour when the electricity price is at the Value of Lost Load (VoLL), or a period when, if demand increases by one unit, load loss will happen in the system (either in another period or in another country) – an alternative but identical definition (ESO, 2022)

[10] Renewable Energy Hub, What Will Dunkelflaute Mean for Renewable Energy Surge, accessed 14.07.2023 https://www.renewableenergyhub.co.uk/blog/what-will-dunkelflaute-mean-for-renewable-energy-surge#:~:text=What%20Is%20Dunkelflaute%3F,each%20November%2C%20December%20and%20January

[11] A brief climatology of dunkelflaute events over and surrounding the North and Baltic Sea Areas, Li, Basu and Watson, 2021 https://www.researchgate.net/publication/355173603_A_Brief_Climatology_of_Dunkelflaute_Events_over_and_Surrounding_the_North_and_Baltic_Sea_Areas

Figure 1: *Frequency of Dunkelflaute for the UK, using different thresholds for capacity factor and time duration, for solar and wind both, in the years 1985 – 2016*



Note: The underlying data, from the Renewables.ninja tool, is generated by simulating all operating wind and solar farms in 'current' locations as of 2016. It investigated how the frequency of Dunkelflaute, using different thresholds for the duration, and the capacity factor (CF) – this is the fraction of wind and solar photovoltaic (PV) power production normalized by the respective installed generation capacity. A 5% CF threshold means that both wind and solar production are at 5% or less of capacity.

Size of events

The risk of large loss of load events is likely to increase. There may be large loss of load events (in absolute terms) simply because of the expected increase in peak demand and hence the total generation capacity in a future system. The potential for large reductions in correlated renewables output may also change the size of events relative to total demand. However, we have not made an assessment of the direction of change.

Volatility of events

In the past, individual system stress events were expected to be relatively similar in nature (i.e. there was expected to be limited variability in the nature of system stress events within a year, and different years had a similar number/type of stress events). In this context, an average measure of reliability adequacy (such as LOLE) was sufficient to target these events.

In the future, system stress events are expected to become more volatile across all three of the dimensions mentioned above (frequency, duration, and size). For example, modelling from

ESO (2022) study suggests that the GB system will be more susceptible to low-probability, high-impact events, meaning that *“in many years, no tight periods on the GB system would be expected, but occasionally in other years, there could be prolonged tight periods that are more challenging”.*

2.2. Challenges with the current approach to setting reliability standards

In this section we set out our findings from the literature relating to whether LOLE continues to be an appropriate metric. This section is structured as follows:

- First, we set out the current approach to LOLE, and the trade-off inherent in the metric;
- Then, we consider whether LOLE continues to be appropriate given the changing nature of system stress events. We explain which aspects of system stress events the LOLE is likely to miss, and why these matter.

Overview of current approach to LOLE

It is impossible to guarantee 100% electricity system reliability, and no system sets this as a target. Unplanned power plant outages (or lower than expected wind or water availability) are, by their nature, random, and can coincide with one another. Even in a system with very large amounts of excess capacity it would still be possible (just extremely unlikely) that coincident failures cause demand to exceed supply. The question for policymakers is therefore how to determine the amount of capacity of the system to ensure a certain level of reliability.

Historically, most systems have relied on setting a reliability standard based on the LOLE. As we will go on to discuss, in a traditional power system LOLE is effective at capturing the key risks related to system stress events. As of July 2022, the reliability standard is set as LOLE in all countries with electricity interconnections to the UK[12]: Belgium (3 hours), the Netherlands (4 hours) France (2 hours), and Ireland (8 hours) (ACER, 2022).

The LOLE represents the number of hours/periods per annum In which, over the long-term, it is statistically expected that supply will not meet demand under normal system operation. It provides a measure of the likely frequency of shortage events, though it is not a measure of the actual scale of any shortfall.

In theory, determining the number of LOLE hours to target is based on a trade-off between:

- the value that customers place on reliable electricity supply; and

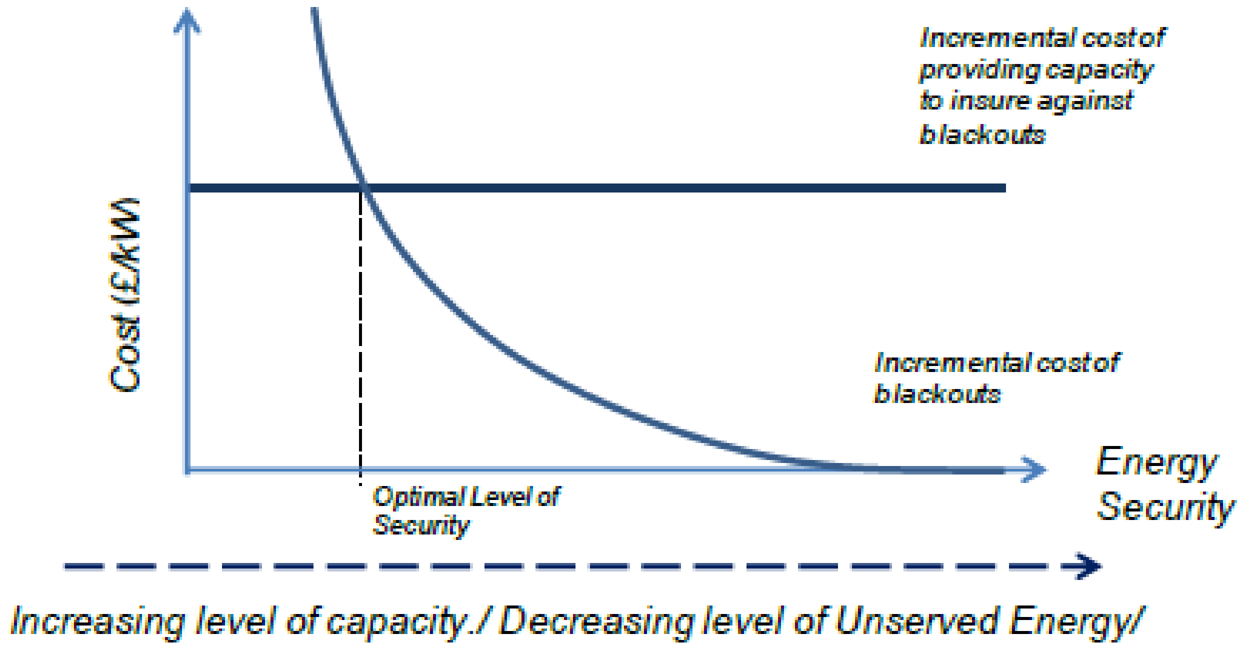
---

[12] Ofgem, accessed 17.8.2023 https://www.ofgem.gov.uk/energy-policy-and-regulation/policy-and-regulatory-programmes/interconnectors

- the cost of adding incremental capacity.

As more capacity is added to a system (i.e. the capacity margin over peak demand is increased), the likelihood of outages is reduced. However, adding capacity has a cost. Moreover, the gains from incremental capacity in terms of reliability diminish with the amount of capacity that is added to the system, such that there is likely to come a point where it is inefficient to keep adding capacity to reduce the likelihood of outages further. This is illustrated in DECC (2013) below.

Figure 2: *Optimal level of security relative to VOLL/CoNE*



The trade-off between reliability and cost means that one approach in determining the level of reliability is to link it to society's willingness to pay for it. Under this approach, the reliability standard is set at the point where the incremental cost of capacity to improve reliability is equal to the incremental benefit to society of improving reliability. A reliability standard set to ensure greater security of supply than this level would imply investing in new capacity where the cost of doing so is greater than the value society is willing to pay for the resulting improvement in reliability.

The ***willingness to pay*** **can be measured by the Value of Lost Load (VoLL)**. This represents the value that customers place on reliability, or their own cost of interruption. There is no one figure for VoLL – it will vary by country (based on macroeconomic conditions), by season, day of the week, time of day, the duration, scale and frequency of the loss of load, and by customer type. However, for the calculation of LOLE, an average value of VoLL is typically estimated. The ***cost of supply*** **can be expressed as the Cost of New Entry (CONE)** i.e. the cost of building new capacity. The LOLE at the optimal level of reliability can be defined as the ratio between these two values:[13]

$$LOLE = \frac{CONE}{VoLL}$$

In GB a LOLE of 3 hours has been adopted since 2014. At the time DECC justified its chosen methodology for calculating the LOLE as follows:

> *“The Value of Lost Load represents the value that customers place on security of supply, or alternatively the cost to customers of being disconnected. The optimal level of security of supply trades off the cost of providing additional capacity against the associated benefit of reduced blackouts that comes with an increase in capacity. This method has the advantage of choosing a level of capacity that is explicitly linked to the value that consumers place on electricity. This should drive a more efficient outcome.” (DECC, 2013)*

A LOLE of 3 hours means that, on average, it should be expected there will be loss of load related to a shortage in supply relative to demand (i.e. not related to network outages) in 3 hours across the year. It is important to note that this does not mean there will be blackouts. In the majority of cases, it is expected that loss of load could be managed by the system operator without significant impacts to consumers i.e. it can take a number of actions that are beyond ‘business as usual’ balancing actions in order to avoid significant load disconnections.

Suitability of LOLE

LOLE can be considered as a relatively opaque measure (ESIG 2021). LOLE only reflects an average number of shortfalls over a particular period. It does not capture any sense of the size of a loss of load event, and its focus on an average value doesn’t reflect anything about the nature of the whole distribution of events. As we will discuss in the next section, similar limitations can be identified for all possible metrics as there is no single perfect metric that captures all dimensions of stress events.

Historically these shortcomings have not been seen as a particular problem. This is principally because individual system stress events were expected to be relatively similar in nature. Policymakers could be reasonably confident as to the size of expected individual events, and therefore by targeting LOLE were implicitly also targeting what was considered to be a reasonable level of expected energy unserved (EEU). However, it would not have been transparent as to what the implicit EEU target would have been.

Going forward, the nature of system stress events may begin to change. Therefore, for a given value of LOLE set as the reliability standard, we may see other dimensions of system stress events begin to change from historic values, e.g. loss of load events may be of a greater size,

---

[13] The derivation of this formula is based on a clear framework set out in the UK context. See Annex C: Reliability Standard Methodology, July 2013, Department of Energy & Climate Change.

or events could become longer, meaning customers experience greater costs. The point was also illustrated by London Economics International who said that while targeting an average LOLE of 3 hours, the magnitude and duration of events is not controlled for (London Economics International, 2023).

It may also be that while the average LOLE is maintained, it could be that the distribution starts to change with respect to LOLE (or other factors) increasing the risk of more extreme events, i.e. the 95th percentile may increase while the average stays the same. The issue of a change in distribution was identified by ESO (2022). To calculate LOLE metrics, the average over all periods in the years and across a large number of simulations is taken (to capture uncertainties in weather and other factors). As the weather becomes more extreme, it may be that, when targeting a LOLE of 3 hours, there is no loss of load in most of the simulations, but with very large outages (in both duration and size) in some scenarios.

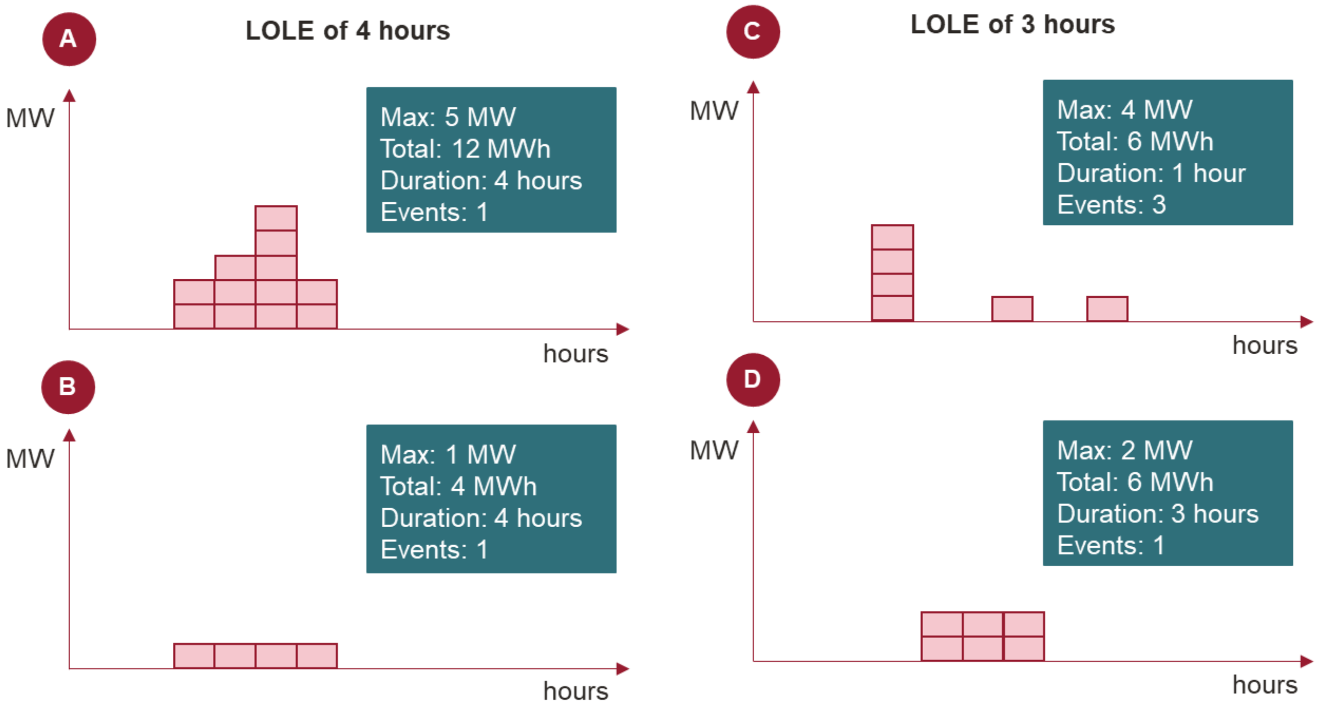
ESIG 2021 provide an illustration (based on a US study) to show how LOLE may not capture the more complex nature of future system stress events.[14] ESIG describe how targeting a single metric can result in a very different experience (and hence costs) for customers. In Figure 3 below we have adapted ESIG's illustration to be more relevant to the GB discussion, however, the broad point being made is the same. From the illustration below the following points can be observed:

- In panels A and B, for a constant LOLE of 4 hours, customers would expect to experience a single event of four hours in duration, but with much greater total size (MWh) and maximum level (MW) in any hour in panel A.
- In panels C and D, for a constant LOLE of 3 hours, customers experience the same total size out outage(s) (MWh), but experience three short events of varying MWs in panel A, and a single long event of constant MWs in panel B.

---

[14] See Figure 7 in ESIG 2021.

Figure 3: *Buildings blocks of resource adequacy*



Source: Frontier Economics, based on ESIG (2021), Figure 7. Each block represents loss of load of a MWh

VOLL may be sensitive to these different dimensions

As the nature of the events changes, the costs imposed on customers of system stress events will also change. In turn, this changes the nature of the trade-off between customer costs of disconnection and the cost of new capacity that underpins the methodology to set LOLE.

London Economics (2013), which was used to set GB's reliability standard, assumed a load-share weighted average across domestic and SME users. It also assumed that the VoLL was the same no matter the size or duration of the loss of load event. This was likely to have been a reasonable assumption in the context of events of short duration and small in size.

However, if events increase in size and duration, a more complex understanding of the relationship between VoLL and the different dimensions of system stress events may need to be developed. As CERRE (2023) write, *"unplanned, large-scale, and long-lasting disruptions of supply services may have a more serious impact"*. Therefore, customers' appetite for risk might also change. After Winter Storm Uri, Vistra Corp, a Texan energy company, write, *"[the storm] effectively chastened the public's and policymaker'' tolerance for load shed events, generally making them more risk averse … [whereas] a reliability standard based on the average outputs from that modelling will be inherently 'risk-neutral'"* (Vistra Corp, 2023).

This implies a non-linear relationship between VoLL and the duration and size of loss of load events. For example:

- Shorter events may be more easily managed by deferring activities for a short period. However, such deferrals become more challenging the longer the event lasts;[15] or
- Smaller events may be more easily managed by disruption to activities with a lower VoLL, whereas larger events will impact on more customers including those with higher values of VoLL.

This has implications for the amount of capacity that is procured to meet the reliability standard as well. For example:

- If for a given value of LOLE, events are now longer, or the risk of more extreme events has increased, it may be the case that the VoLL associated with that level of LOLE has increased relative to historical levels. This would also imply that, for a given value of the cost of new entry, the optimal level of capacity will have also increased; and
- System stress events may now occur outside of peak hours (due to, for example, no solar energy being produced at night), when VoLL is likely to be lower. This would suggest a different trade-off with lower levels of optimal capacity.

2.3. Alternative reliability standard metrics

Figure 4 below shows the reliability standard metrics that we have identified in our literature review. Our review identified a number of metrics that can be used to target different dimensions of system stress events.

However, the literature did not identify a single metric that is superior to the others at targeting the type of system stress events that we expect in the future. We elaborate on the range of different potential metrics which could be used in the last section of this chapter.

In the figure below we have grouped the metrics that we have identified according to the four dimensions of system stress events which we have introduced in Section 2.2. We describe each of the metrics in turn, grouped by dimension.

The metrics that we discuss in this section are probabilistic metrics calculated for the purposes of making planning assumptions related to capacity adequacy. These metrics are probabilistic as they are calculated by:

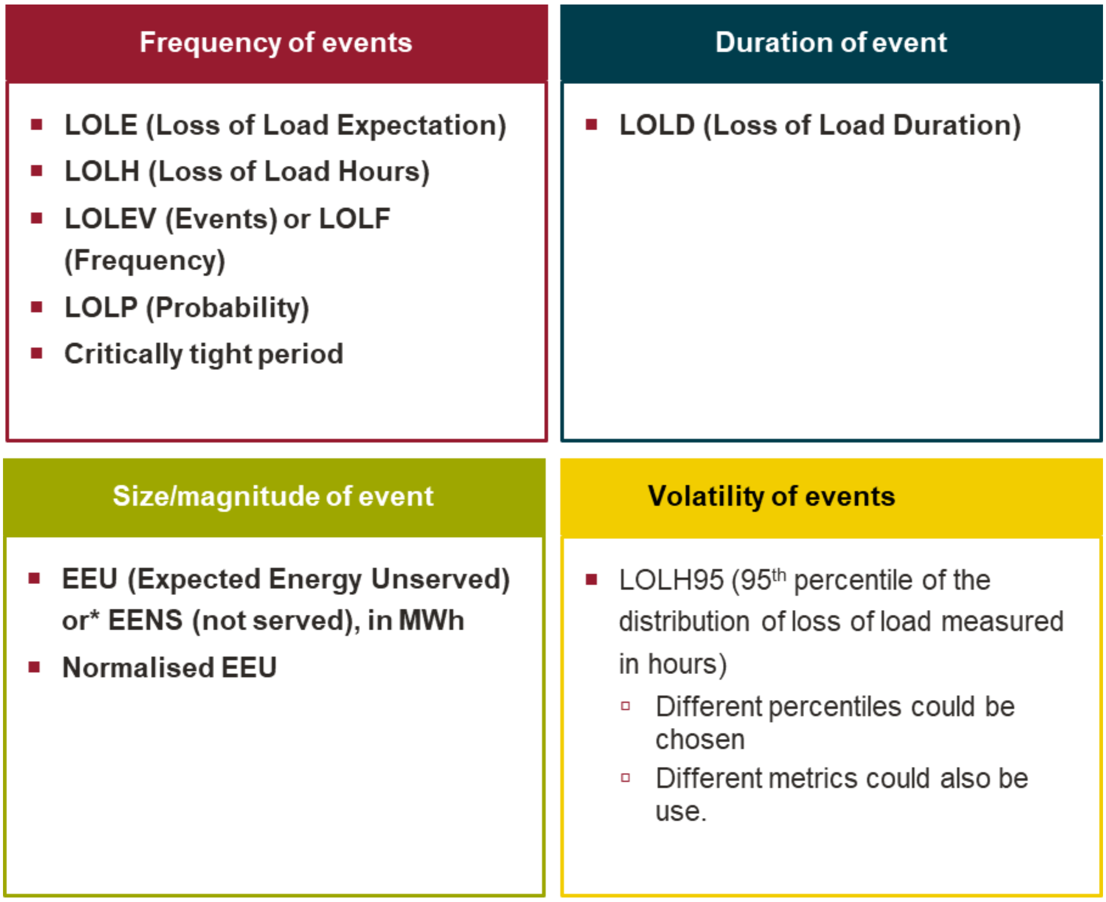
- first, simulating the supply and demand balance for a given year, based on possible values drawn from a distribution for key uncertain drivers (e.g. how cold a winter is, level of demand, how many power plants fail, etc);
- second, for that modelled run calculating a summary statistic depending on the reliability metric of choice (e.g. number of outages, or average duration of outages in the year); and
- third, repeating this process many times (typically thousands of runs are carried out) and then calculating an average of the summary statistic across all of the modelled runs.

---

[15] In contrast, some literature related to VoLL says that for some customers the first hour is the most costly given the surprise of the outage, after which VoLL declines when they are able to readjust their behaviour.

Therefore, these metrics are model-generated expected values used to guide the expected future performance of the electricity system. We do not consider measures of actual performance of the system based on historical data given they are not relevant for planning purposes. [16]

Figure 4: *List of reliability standard metrics according to the four dimensions of system stress events*



Frequency of events

- LOLE (Loss of Load Expectation)
- LOLH (Loss of Load Hours)
- LOLEV (Events) or LOLF (Frequency)
- LOLP (Probability)
- Critically tight period

Duration of event

- LOLD (Loss of Load Duration)

Size/magnitude of event

- EEU (Expected Energy Unserved) or* EENS (not served), in MWh
- Normalised EEU

Volatility of events

- LOLH95 (95th percentile of the distribution of loss of load measured in hours)
    - Different percentiles could be chosen
    - Different metrics could also be use.

*Note: *Listing metrics using "or" does not mean they are identical (although they might be), but that they are similar enough to be considered together*

Frequency

There are a number of metrics that are designed to measure the frequency of system stress events. These metrics are based on counting the number of stress events where a loss of load is expected to occur over a period of time. Stress events are usually defined in increments of hours or days, but other definitions can be used (e.g. continuous hours could count as 1 event).

---

[16] Examples include: 1) AIT (Average interruption time) which is a measure for the amount of time the supply is interrupted, expressed as the total number of minutes that the power supply is interrupted during the year 2) SAIDI (System average interruption duration index) which is the probability of involuntary interruption from the point of view of each individual customer (whereas LOLP/LOLE describe the probability some consumers will involuntarily not be supplied)

In the following sections we describe a number of frequency metrics and explain how they are usually calculated.

LOLE (loss of load expectation)

In general, a LOLE measure relates to the expected number of events in a study horizon period, whether that event is an hour or a day.

The definition of LOLE used in much of the literature (which is largely from the US) is based on the expected number of days per year in which an event for which available generation capacity is insufficient to serve the daily peak demand is expected to happen. This is the definition applied in a number of US jurisdictions (NERC 2016). In some markets, the assessment looks over all hours of the day, and therefore, a LOLE of 1 day a year could imply an event of 1 hour or more during the day.

In GB, LOLE is defined as the expected number of hours per year in which supply is not expected to meet demand, i.e. the total number of hours for which EEU (see below) is greater than zero. Given GB's definition, its measure of LOLE is the same as the commonly accepted definition of loss of load hours LOLH, i.e. the expected number of hours per year when a system's hourly demand is projected to exceed the generating capacity.

LOLE is used as a reliability standard metric in many of countries. For example, GB uses a LOLE of 3 hours per year along with many other European countries. Many parts of the US cap LOLE at 0.1 days per year (London Economics International, 2023).

Singapore caps LOLH at 3 hours per year. It moved from LOLP (see below) to LOLH in 2017, *"to better measure the hourly impact of intermittent generation resources"* (London Economics International, 2023).

LOLP (loss of load probability)

LOLP is defined as the probability of system daily peak or hourly demand exceeding the available generating capacity during a given period (Heylen et al., 2018). The probability can be calculated using only daily peak loads or all the hourly loads in a given period (NERC, 2016). The following relationship between LOLP and LOLE holds:

$$LOLP = LOLE / N.$$

Where N is the number of possible events in a year (365 if LOLE is measured in days, 8760 if LOLE is measured in hours). In the GB context, LOLP is implicitly 3 / 8760 hrs = 0.034%

A number of countries set their reliability standards based on LOLP. For example, Mexico caps LOLP in two ways, where maximum LOLP (based on the government's reliability policy) is capped at 0.2178%, and efficient LOLP (developed to reflect a comprehensive metric adjusted for the Mexican context) is capped at 0.0315%.

LOLF (loss of load frequency)

LOLF is the number of events in a year where there is insufficient supply available in the market. Events can last for one hour or for several continuous hours.

The following relationship between LOLE and LOLF holds:

$$LOLE = LOLF * LOLD$$

Where LOLD is the loss of load duration (see below).

A similar metric is **LOLEV (Loss of load events).** This is the number of events in which some system load is not served in a given year (Heylen et al., 2018) (NERC, 2016).

Critically tight periods

A critically tight period is defined as an hour when the electricity price is at the VoLL. This metric was used by ESO 2022 to identify when demand was met but the system was very close to loss of load. In particular, ESO 2022 defined ‘critically tight periods’ as periods when, if demand increases by one unit, loss of load would occur in the system. Under the ESO definition, with an efficiently operating wholesale market prices would be expected to be at VoLL in such a situation.

Duration

Duration reliability metrics are designed to target the length of a stress event. The length can be measured in hours or days.

The most relevant metric in this group is the **LOLD (loss of load duration).** LOLD is defined as the *average* length of an event in a year where there is insufficient supply available in the market. It is useful to note that LOLD can be defined as LOLE divided by LOLF.

$$LOLD = LOLE / LOLF.$$

How are LOLE, LOLD and LOLF calculated?

Each of these metrics is calculated as the average of a metric across a range of simulation of future system which generates a distribution of outage events. The difference between LOLE, LOLD, and LOLF is in the metric calculated for each simulation:

- LOLE is calculated as the average of the **number of hours** in which supply is not expected to meet demand in a year, across the simulations
- LOLD is calculated as the average **length** of outage events, across the simulations
- LOLF is calculated as the average **number** of outage events in a year, across the simulations.

Given the relationship between LOLE, LOLD and LOLF, the LOLE distribution can be broken down into the LOLD and LOLF distributions (and vice versa).

Size

These reliability metrics target the magnitude of a stress event. The most common size reliability metrics that we have identified are based on the **Expected Energy Unserved (EEU)**.[17] EEU is defined as the amount of electricity demand measured in MWh that is expected not to be met by generation in a given year. This combines both the likelihood of any potential shortfall, with the potential size of any shortfall (ESO, 2022).

Some countries use EEU as a reliability standard. For example, Australia caps the EEU at 0.002% of actual annual operational demand (London Economics International, 2023).

One of the main advantages of using such a metric is that it places a higher weight on large and disruptive system stress events, rather than treating each loss of load event equally as is the case with LOLE, LOLH and LOLP. If combined with VoLL, an estimate of the cost of the expected outages can also be obtained.

It is possible to transform measures of size based on EEU into relative measures, which allows for a comparison of relative risk levels across systems . The choice of normalisation can be applied to both EEU itself, and/or a tail measure (see the section on volatility below).

For example, EEU may be normalised as follows:

- EEU relative to **peak demand**: Brazil caps EEU at the 95th percentile relative to the monthly peak demand
- EEU relative to **total demand**: Australia’s EEU is capped at 0.002% of total demand annually; Brazil’s EEU is capped at the 99th percentile relative to annual demand. ERCOT (2014) mentions a normalised EEU (%) measuring the total annual MWh of firm energy expected to be shed, divided by the total MWh of annual system load, to represent the percentage of system load that cannot be served due to supply shortages.

It is important to note that these transformations may impact optimal decisions over the medium term, given that reliability standards are likely to be in place over a number of years. Over time, total demand may develop differently to peak demand. Therefore, one metric may imply a greater or lower capacity requirement than another.

Volatility of system stress events

Traditional reliability metrics target expected or average parameters. These metrics are more appropriate when system stress events are relatively homogenous (i.e. in a given year stress events are similar and over different years stress events are expected to be similar in terms of number/type of stress events) and therefore an average would capture well the nature of the expected system stress events.

As mentioned above, in the future system stress events may become less homogenous. In order to target these events other statistics of the distribution of the metrics presented in the

---

[17] Sometimes referred to as Expected Unserved Energy (EUE), or Expected Energy Not Served (EENS).

previous section could be (and have been) considered. For example, rather than defining the standard based on average/expected values, it is possible to define the standard based on (say) the 95th percentile of the distribution. In this case, the 95th percentile represents the maximum value that the metric under consideration (say LOLE) should achieve during severe conditions which occur with a probability of 5% (e.g. a very cold winter that occurs 1 in 20 years).

Some countries have already considered these metrics that capture different aspects of the distribution, including Belgium and Brazil (in combination with traditional metrics):

- Belgium uses a two-part reliability criterion on LOLH which takes into account both an average LOLH criterion (3 hours/year) and an LOLH95 criterion (20 hours/year), which makes sure there are enough resources to cover a statistically abnormal year (ESIG, 2021).[18] In other words, Belgium not only targets an average value of LOLH, but it also wants to build sufficient capacity to ensure that very rare but more extreme events are not expected to exceed 20 hours.
- Following a drought in 2015, Brazil adopted in 2019 a reliability standard based on multiple criteria that target tail risk (rather than average values). It is worth nothing that Brazil has a hydro-based system, which is different from GB's thermal based system, and therefore the key issue for Brazil relates energy constraints rather than capacity constraints. The regulator produces new supply based on a number of criteria that depends on the tail of the distribution (taken from London Economics International, 2023):
    - If the monthly marginal operating cost under a 10% worst case scenario exceeds a threshold value, new supply is needed;
    - On a monthly basis the EEU under the 5% worst case scenario cannot exceed 5% of monthly peak demand – this is illustrated in the chart below;
    - On an annual basis, the EEU under the 1% worst case scenario cannot exceed 5% of energy demand
    - It also caps LOLP at 5% (for unserved load of any magnitude)

Recently, London Economics International (2023) explains that reliability standards that are based on averages (such as LOLE) are risk neutral, "*meaning that they value every unit of unserved energy or hour with unserved energy in the probability distribution equally*". However, "*consumers are not risk neutral*". This applies particularly in the wake of climate change driving more extreme weather events: as Vistra Corp write, Winter Storm Uri in 2021, which led to widespread blackouts, "*generally ma[de] them* [the public and policymakers] *more risk averse*" (Vistra Corp, 2023).

To illustrate this point, the LOLE standard only reflects an average number of shortfalls, and does not reflect anything further about the distribution of events. This means that a LOLE standard of 3 hours can be achieved when either 1) loss of load is 3 hours in all modelled scenarios; or 2) loss of load is 60 hours in 5% of modelled scenarios, and 0 hours in the

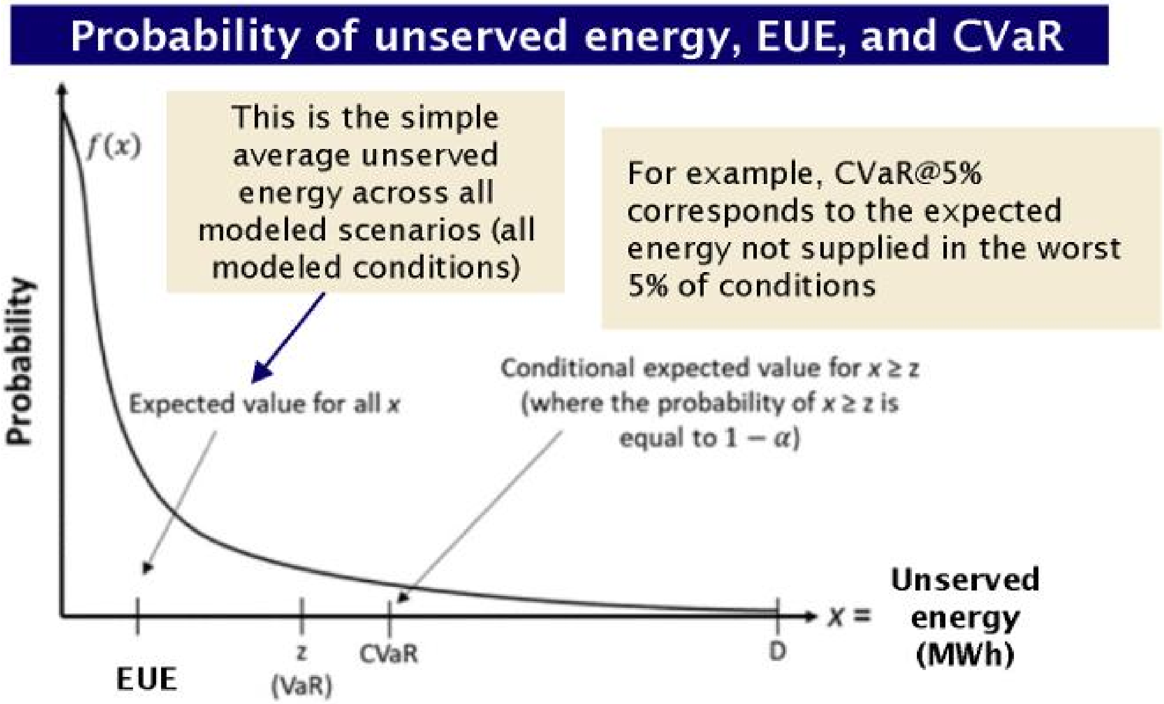
---

[18] See also the European Commission's decision on state aid for Belgium, Section 2.3.1 (15). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022D0639&qid=1652500553540

remaining 95% of scenarios (a risk neutral approach, as 3 = 60*0.05 + 0 * 0.95). However, the loss of load in 60 hours across one year under 2), may be disproportionately more costly for customers than the outages under 1).

Consequently, London Economics International recommended that the Public Utility Commission of Texas should consider metrics that focus on tail risk, or outage risk under extreme weather conditions (London Economics International, 2023). This approach was also supported by Vistra Corp (Vistra Corp, 2023), who wrote that "*Vistra continues to recommend … a measure to take into account the "risk-averse" nature of reliability events, such as the 'Conditional Value at risk' (or CVaR) framework or something similar to it*". The CvaR is shown in Figure 5.

Figure 5: *Probability distribution of unserved energy, EUE*, and Conditional Value at Risk (CvaR)**



Source: London Economics International (2023)

*Note: *EUE stands of Expected Unserved Energy, this is equivalent to EEU.*

***CvaR@X% stands for Conditional Value at Risk. It describes the thickness of the tail of reliability events defines as the man value of the X% worst cases;*

2.4. Conclusions

In this section we summarise our conclusions based on the findings from the literature we have presented above. We find that:

- As system stress events become more complex, it may be more appropriate to target a combination of metrics;
- The literature does not provide much clarity on what is the optimal combination nor on how to set target levels using multiple metrics;
- We note that empirical analysis (including that carried out as part of this study) could help in deciding on the optimal combination of metrics, given that this will depend on relevant scenarios for how the system is expected to evolve in GB.

We elaborate on our conclusions in the sections below.

As mentioned previously, despite extensive literature on reliability standards in general, we have only identified relatively few studies tackling how reliability standards may need to evolve as the electricity system changes in future. Additionally, as it currently there are very few countries that have tried to adjust their reliability standards to address these challenges. While our review has been carried out quickly, we think it is clear that further development of the literature will be required in this area. We suggest possible extensions below.

A combination of metrics could be better at targeting more complex system stress events

Our literature review has identified a broad range of metrics that can be used to target different dimensions of future system stress events. Most metrics only target a single dimension of system stress events. Historically, this has not been a problem. As described in the first sections of this chapter expected loss of load events were quite homogenous, short in duration and relatively limited in size. Because of this a single metric was found to be sufficient to set reliability standards.

In the future, we expect system stress events to become more complex, with different dimensions of system stress events likely to become more important even if the same LOLE level is being targeted. The literature review indicated that there could be value in targeting more than one metric, rather than developing a single superior metric. However, the literature does not provide a clear indication as to what that combination should be.

A relevant example is the recent consultation by the Public Utility Commission (PUC) of Texas, which is developing a reliability standard for the Texas power region ERCOT. PUC 2023 summarised the responses to the consultation as follows *“[w]hile there was some support to keep the reliability standard simple, there was a general acknowledgement that using a single reliability metric would be insufficient, and that there was value in pursuing a multi-metric approach.”* London Economics International (2023) in particular recommended using *“a combination of an expected outage metric such as EEU or LOLH and a tail risk-based metric”*.

As mentioned earlier, some jurisdictions already use a combination of metrics. For example, Belgium has implemented a multi-metric approach. Belgium uses a two-part reliability criterion on LOLH which takes into account both an average LOLH criterion (3 hours/year) and LOLH for the 95th percentile of the distribution of 20 hours/year.

In a separate report, ESIG (2021) recommend *reporting* more metrics: they write that as “grids around the world continue to decarbonize and integrate renewable energy”, considerable work is needed to fully define what robust resource adequacy looks like. They suggest that resource adequacy analysis can be improved by reporting a broader range of resource adequacy metrics than simply an average LOLE, including hourly EEU and additional information on the distribution of outages.

There are some practical challenges in targeting a combination of metrics

In practice, targeting more than one metric implies sufficient capacity is required such that each metric target level is met. This could mean that either:

- *Additional capacity* is required to secure the system consistent with the new metric as well as the LOLE target, and as a result, an average value below the actual LOLE target could be achieved; or
- *No additional capacity* is required on the grounds that achieving the average LOLE target also achieves the additional target.

While there appears to be a reasonably clear argument for considering multiple metrics, the literature does not provide a clear analytical framework for setting the target level of multiple metrics that strikes an optimal balance between the value of lost load and the cost of incremental additions to the capacity mix. For example, we have not been able to ascertain from public documents the basis on which Belgium set its two targets.

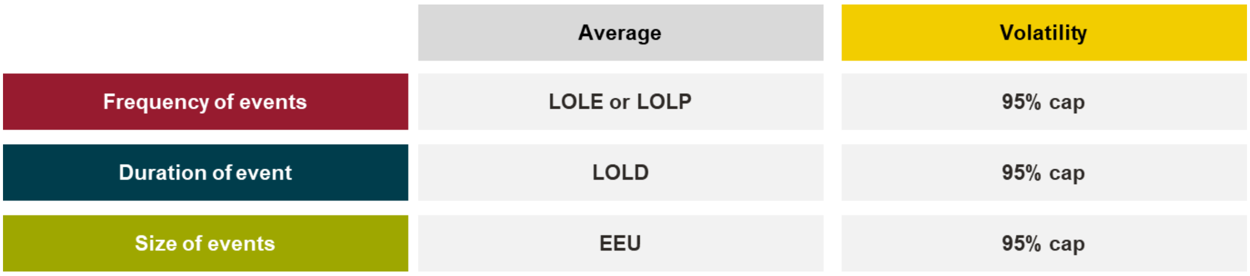
As a result, going forward there may be a need to build on the analytical framework set out when DECC set its reliability standard in 2013 to consider the implications of different dimensions of system stress events, and the non-linear relationship between some of the dimensions and VoLL, e.g. the fact that VoLL may increase with events of longer duration, or the fact that deeper events are proportionately more costly than shallower events. This is clearly a complex area of theoretical analysis, but there may be feasible extensions that could be made which would help bring more structure to decisions on the appropriate target levels.

Empirical analysis can help in deciding the optimal combination of metrics

In practice, countries are likely to want to target those metrics which are most relevant to their system. While in theory we have demonstrated that there may be good reason that particular metrics across each of these dimensions may become important, it is ultimately an empirical question based on relevant scenarios for how the system is expected to evolve.

In the next sections of this report, we explore the empirical question in the GB context to identify possible relevant additional characteristics of system stress events. Based on the dimensions discussed in this review, a number of metrics could be targeted as illustrated in the table below. For example, it would be possible to consider an average metric and/or a tail (“volatility”) measure of frequency, duration, and size.

Figure 5: *Potential metrics that could be target in the GB context*



	Average	**Volatility**
Frequency of events	LOLE or LOLP	95% cap
Duration of event	LOLD	95% cap
Size of events	EEU	95% cap

If any metrics are judged to be relevant in GB, then the next question would be at what level should they be set. Absent an analytical framework based on the trade-off between VoLL and CONE, then one approach would be to set any additional metric at the level observed in current modelling when determining the level of capacity consistent with the LOLE target of 3 hours. Implicitly this assumes that the value associated with the additional metric was consistent with the trade-off being made between VoLL and CONE when setting the LOLE. We explore these possibilities in the next sections.

Further considerations

One final consideration in relation to setting a reliability standard relates to coordination of possible changes in approach with neighbouring markets. Currently, there is a broad consistency across many European countries of setting a LOLE standard of 3 hours, though we note that Ireland has a standard of 8 hours. There is value in consistency, as significant differences can lead to a risk that countries free-ride on a neighbour’s security of supply.

If we imagine a system with minimal interconnector constraints, reliability is effectively shared regionally such that a country with a higher LOLE standard (i.e. a lower standard of reliability) compared to its neighbours would be able to free-ride on the higher level of reliability of its neighbours. For example, if there is surplus of available generation in one part of the region (which is made more likely due to the higher standard of reliability of some countries in the region) but unmet demand in another, the surplus will naturally contribute to meeting the unmet demand in the country with a lower standard of reliability. In this sense, a country can declare a higher LOLE standard, but they may in effect receive a standard related to the lower LOLE standard (i.e. greater reliability) in neighbouring systems.[19]

---

[19] To the extent that the UK target set out in the Energy White Paper, 2020, to “*realise at least 18GW of interconnector capacity by 2030*” reduces periods of constraints over interconnectors, then this effect may increase in relevance for the UK.

3. Modelling approach

3.1. Modelling process and metrics tested

For the second stage of the study, modelling has been carried to assess how the characteristics of system stress events change over time in a transitioning system, and then test the implications of using different reliability standard metrics. This is done using LCP Delta's Unserved Energy Model (UEM), which is currently used by National Grid ESO in its role as EMR delivery body. When running the UEM, over 1000 simulations are modelled to capture variations in intermittent generation, demand and plant outages. For each scenario, the UEM has been run from 2025 to 2034 with the modelling split into two parts:

- First a counterfactual is modelled with the reliability standard metric matching the current metric; LOLE of 3 hours. Within this counterfactual, we assess how the value of each reliability standard metric changes over time across each scenario.
- Secondly, the existing reliability standard and an alternative reliability standard metric have been modelled in order to understand the impact that a combined reliability standard metric could have on the level of capacity that needs to be procured. The reliability standard metrics modelled have been chosen based on findings from the literature and the results from the first part of the modelling. In choosing the target level for the alternative metrics, it is assumed that these are maintained at the current (i.e. 2025) level as it is assumed that the current level of these metrics is the desirable level.

Based on the findings from the literature review, a list of metrics was identified to review as possible reliability standard metrics. The metrics chosen cover the 4 key areas identified for security of supply events; frequency, size, duration and volatility. For each of the metrics, the model allows us to explore both the average and the distribution across all simulations modelled. In particular, the 95th percentile was extracted for each metric to capture impacts on volatility. The metrics modelled are:

- **Loss of Load Expectation (LOLE)** – the average number of hours per year in which it is expected statistically that supply will not meet demand.
- **Loss of Load Frequency (LOLF)** – the number of events in a year where there is insufficient supply to meet demand.
- **Loss of Load Duration (LOLD)** – the average length of an event in a year where there is insufficient supply to meet demand.
- **Expected Energy Unserved (EEU)** – the amount of electricity demand – measured in MWh – that is expected not to be met by supply in a given year.
- **Normalised EEU** – EEU normalised based on peak demand.
- **Critically Tight Events** – the number of hours in a year for which it is expected that demand is over 99% of the total available supply.

The following metrics have not been taken forward into the modelling:

- **Loss of Load Probability (LOLP**) describes the probability of at least one system security event occurring in a given time period (e.g. over a year). However, it does not capture the characteristics of system stress events like their frequency or duration.
- **Average Interruption Time (AIT)** and **System Average Interruption Duration Index (SAIDI)** both relate to the actual interruptions as experienced by consumers rather than system stress events that can often be mitigated without intervention by the system operator. The actual pattern of interruptions is highly dependent on the system operator's capabilities and locational aspects at the time. As such, these metrics are typically used as a statistically descriptive metric post-event and include network-related outages as well.

3.2. LCP's Unserved Energy Model (UEM)

For this work, the primary modelling tool used was LCP's Unserved Energy Model (UEM). This is LCP's stochastic security of supply analysis tool, and is used by:

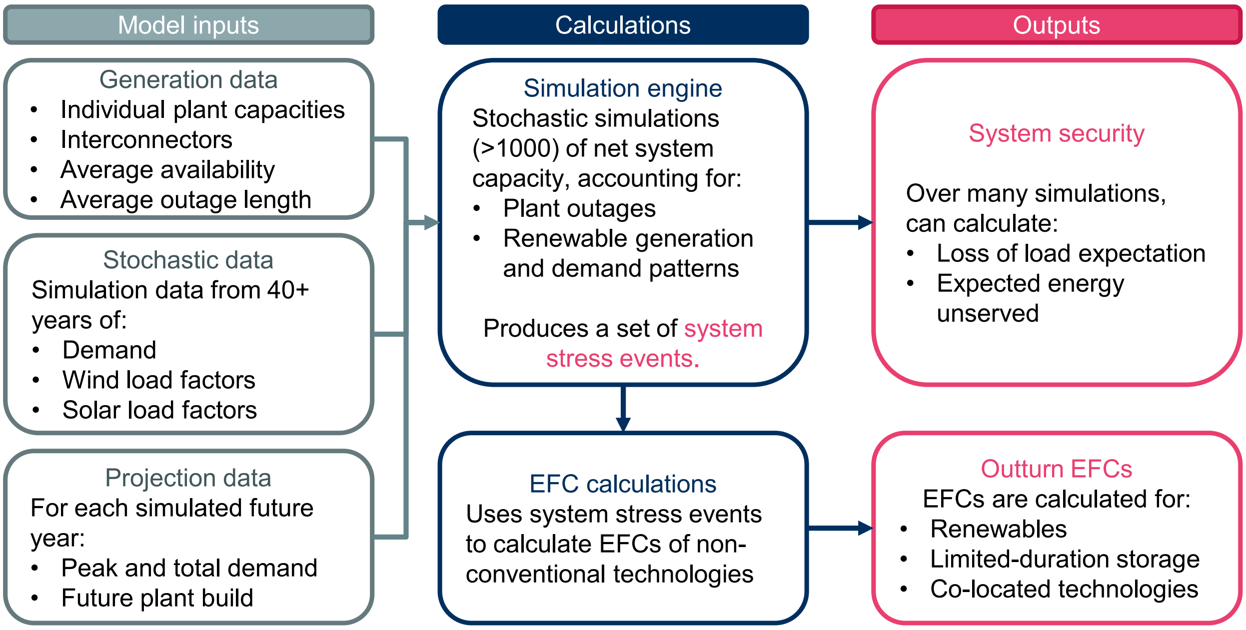
- National Grid ESO to set the derated capacity requirement for Capacity Market auctions, targeting a Loss of Load Expectation of 3 hours in GB[20].
- DESNZ as a component of the Dynamic Dispatch Model (DDM). Within the DDM, the UEM is used to calculate the derated capacity requirement in future CM auctions, which informs the investment decisions made within the DDM.
- National Grid ESO to calculate derating factors for onshore wind, offshore wind, solar and limited-duration storage.

The diagram below gives a high-level overview of the UEM methodology for GB system security.

---

[20] NGESO use the UEM within the DDM to calculate the requirement and standalone UEM for derating factors

Figure 6: *Overview of LCP's Unserved Energy Model (UEM)*



To capture key uncertainties, over 1000 simulations are modelled for each year and scenario. These simulations capture variation in intermittent generation (wind and solar output), demand, and plant outages. For every simulation, the key reliability metrics are outputted allowing the average and the distribution of these metrics across the simulations to be explored. The number of simulations were chosen such that these metrics would converge to a single value, so further simulations would bring little additional value.

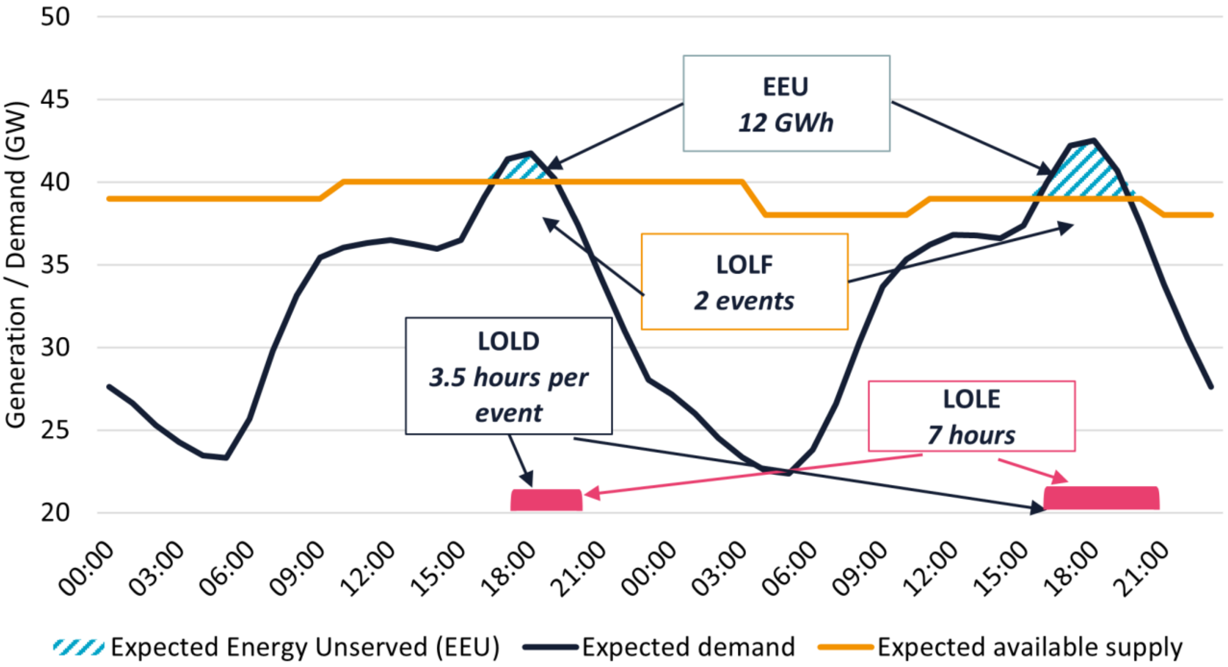
The UEM has two run modes: time-collapsed and sequential. The time collapsed methodology allows for single point estimates of LOLE and EEU based on probabilistic calculations that combine the distributions for intermittent generation, demand, and conventional plant availability (conventional technologies are defined as non-intermittent in this study). The “sequential” mode uses a Monte Carlo (MC) approach. In each MC simulation the year/winter is simulated at half-hourly granularity. Each half-hour is simulated sequentially to capture the hour-to-hour dependencies (length of outages, wind patterns, demand shape).

We use the sequential mode to output the actual simulated system stress events and therefore the distributions of the frequency and duration of events across all simulations can be plotted. This approach also allows the calculation of the summary metrics of the frequency (LOLF) and duration (LOLD) of system stress events as well as critically tight events.

Calculating Security of Supply metrics

As outlined above, there are various metrics that can be used to measure security of supply. The metrics modelled (as outlined in 3.1) are outputted from the UEM allowing for a detailed analysis of security of supply metrics and the ability to explore alternative reliability standards. The key metrics are outlined in the diagram below which shows how different metrics are derived over two example days for two system stress events.

Figure 7: *Security of Supply metrics over two example days*



The below outlines how each metric is calculated within the UEM in more detail.

- The LOLE for a particular year under a given scenario is estimated by calculating the number of hours unserved for each simulation as described above and then averaging across all simulations. LOLE is made up of two parts, frequency and duration.

$$LOLE = \frac{\sum_{i=1}^{NumberOfSims} NumberOfHoursUnserved_i}{NumberOfSims}$$

- The EEU for a particular year under a given scenario is estimated by calculating the total unserved energy for each simulation as described above and then averaging across all simulations. EEU can also be normalised based on peak demand.

$$EEU = \frac{\sum_{i=1}^{NumberOfSims} TotalUnservedEnergy_i}{NumberOfSims}$$

- The LOLF is the number of events in a year where there is insufficient supply available in the market. This forms one part of the LOLE.

$$LOLF = \frac{\sum_{i=1}^{NumberOfSims} NumberOfEventsUnserved_i}{NumberOfSims}$$

- The LOLD is the average length of an event in a year where there is insufficient supply available in the market. This forms the other part of the LOLE

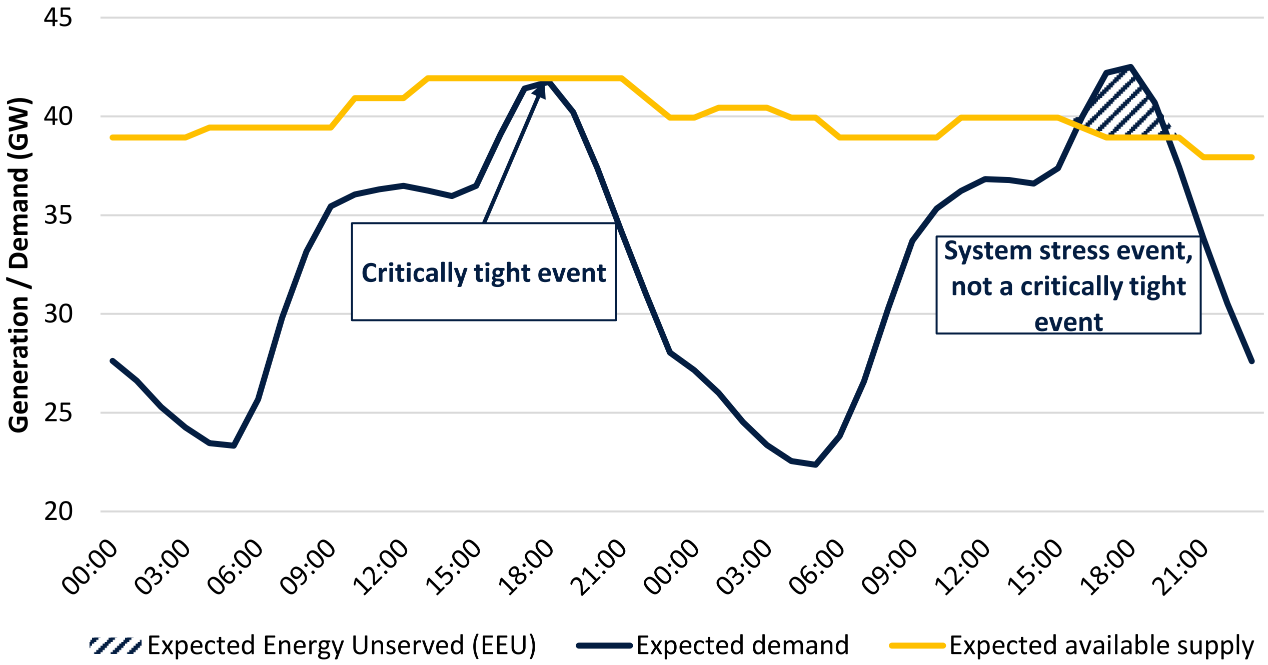
$$LOLD = \frac{\sum_{i=1}^{NumberOfSims} NumberOfHoursOfUnserved_i}{\sum_{i=1}^{NumberOfSims} NumberOfEventsUnserved_i}$$

- Based on the findings from the literature review, we also include a metric for critically tight events. In this study, a critically tight event is defined as an event where demand is over 99% of the total available supply. Previous studies have used different, but similar definitions (ESO, 2022 used a definition where the loss of an additional unit of generation would trigger a system stress event). In practice, a time period is likely to be regarded by the markets and by broader society as 'critically tight' based on the margin between the total available supply and the demand, with wholesale prices expected to approach VoLL as this margin decreases and the probability of lost load increases. The key issue in defining critically tight events is the associated costs to the consumer (both direct and indirect) and the definition used in this study is consistent with this definition.

  An illustrative example can be seen in the diagram below. Similar to the other metrics, the model outputs the average total length of critically tight events in a year:

$$Avg\ total\ critically\ tight\ length = \frac{\sum_{i=1}^{NumberOfSims} NumberOfHoursCriticallyTight_i}{NumberOfSims}$$

Figure 8: *Example of Critically Tight Event*



Calculating the capacity target

The UEM also outputs the target capacity (i.e. the de-rated capacity requirement) based on the desired reliability standard. The methodology for calculating this is described below using LOLE as an example but other reliability standard metrics can also be used.

In the sequential mode, the value of LOLE can be calculated , and this can be used to determine the target capacity. The target capacity is calculated by incrementally adding or subtracting firm capacity to an initial fleet (which should represent a best estimate of the actual mix of technologies that will be online for the target year), and calculating the LOLE at each step . The target capacity is then the value corresponding to the LOLE target.

The LOLE is calculated after each addition of firm capacity in the sequential mode in the UEM by taking the average of the length of system events across all simulations.

Calculating Equivalent Firm Capacity (EFC)

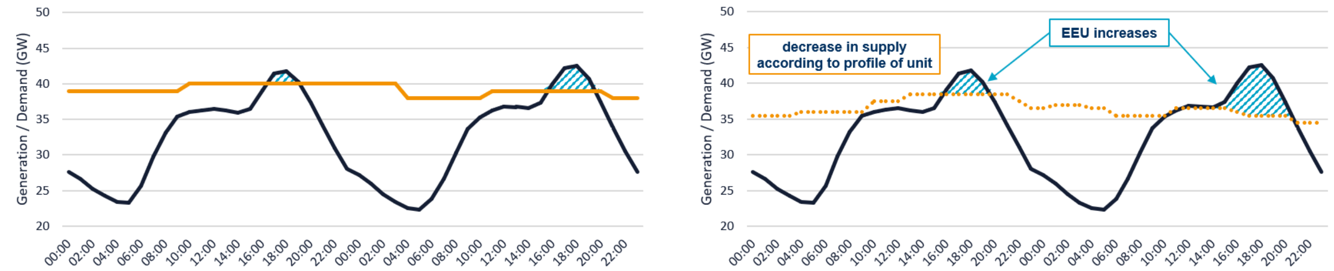
A key output from the UEM and relevant to this study is the Equivalent Firm Capacity (EFC) of different technologies. The Equivalent Firm Capacity (EFC) estimates the contribution of a certain technology to security of supply and is used as the basis for the derating factors used in capacity market auctions by National Grid ESO for intermittent renewables, such as wind and solar, and limited-duration storage technologies. EFCs are a useful output to consider in this study as changing the reliability standard could impact the EFCs for different technologies. More detail on how EFCs are used in capacity market auctions can be found in NG ESO's Electricity Capacity Reports (ECR).[21]

For a given unit(s), the EFC is defined as the amount of firm capacity required to replace the unit(s) and result in the same EEU. In other words, the model calculates the amount of "firm" (100% reliable) capacity that would need to be added to achieve the same level of security if the unit in question was removed. An illustrative example of calculating the EFC for a particular unit is outlined below.

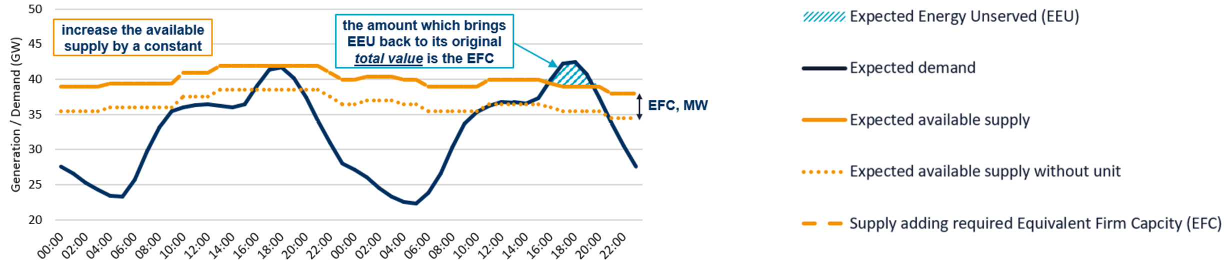
Figure 9: *Summary of the EFC Methodology used in the UEM*

1 the ***EEU*** is the MWh amount of electricity ***demand*** across a period that is not expected to be met by ***available generation***

2 the ***EEU*** may increase if a generating unit or interconnector previously expected to contribute to supply in a tight period is **assumed to be unavailable**



3 the constant amount of ***extra supply*** (i.e. ***firm capacity***) required to reduce the ***EEU*** back to its original total is the Equivalent Firm Capacity of the unit



---

[21] EMR Portal - Capacity (emrdeliverybody.com)

3.3. Modelling limitations

Within the modelling, there are some key limitations and assumptions. The key limitations to be aware of:

- The Unserved Energy Model captures variations in inputs based on historic data on weather (wind output, solar output, temperature's impact on demand), electricity demand and power plant outages. Any risk due to other factors would not be captured, and if the variability profile of these factors changes in the future (e.g.: changing weather due to climate change), this would also not be fully captured, potentially under-estimating the modelled volatility of some of the metrics presented in this report. In addition, a scenario-based approach does not fully consider the risks inherent in the build-out and retirement rates of capacity within a particular scenario.
- Some aspects of system event management will require policy decisions that are yet to be made. Critically tight periods are rare and there has not been a system stress event since the introduction of the Capacity Market, therefore it is difficult to gauge how different plant types, including limited-duration storage plants, will behave in these time periods and reflect this in the modelling.
- To arrive at the exact level of LOLE and LOLD that was specified in a given scenario, the model shifts the level of firm capacity added to the total available supply in the main case. NGESO uses the same approach when they run the UEM to derive target capacities for the Capacity Market auctions and we are using shifts that are comparably sized. However, in practice individual power plants would be added or removed to the system to meet the target security of supply. The type of power plant that are added and removed would have a small impact on the outcome (due to their forced outage rate) – this would be different in different scenarios and would likely change as the power system evolves and the market environment changes.

3.4. Scenarios

Different market scenarios have different assumptions about the makeup and overall level of supply and demand varies in the future, potentially leading to different outcomes when it comes to security of supply. Therefore it is prudent to explore the impacts on reliability standards across different scenarios. To ensure a wide range of uncertainties are captured, two scenarios have been used within the modelling: the DESNZ Net Zero Higher Demand Scenario and the National Grid ESO System Transformation scenarios from FES 2022[22].

The GB capacity mix and demand assumptions are based on the DESNZ Net Zero Higher Demand (NZH) scenario as provided by DESNZ. This scenario was produced using the Dynamic Dispatch Model (DDM)[23], which is provided by LCP to DESNZ and published in

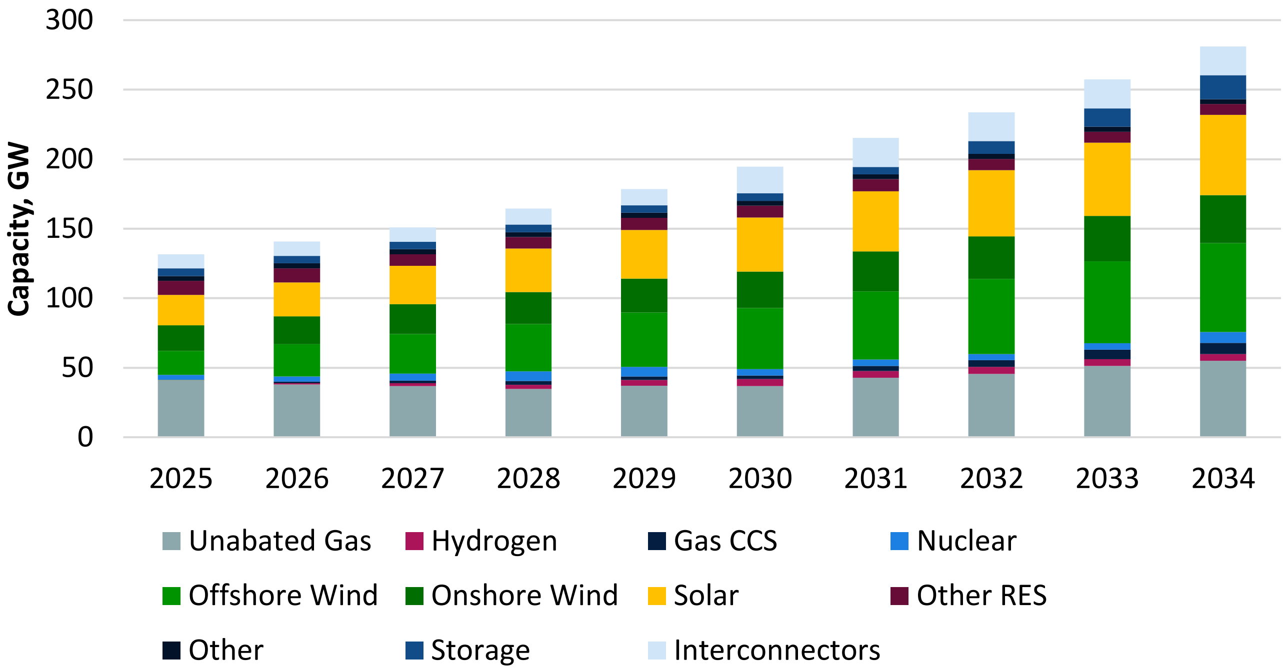
---

[22] https://www.nationalgrideso.com/document/263951/download

[23] Dynamic Dispatch Model (DDM) - May 2012 - GOV.UK (www.gov.uk)

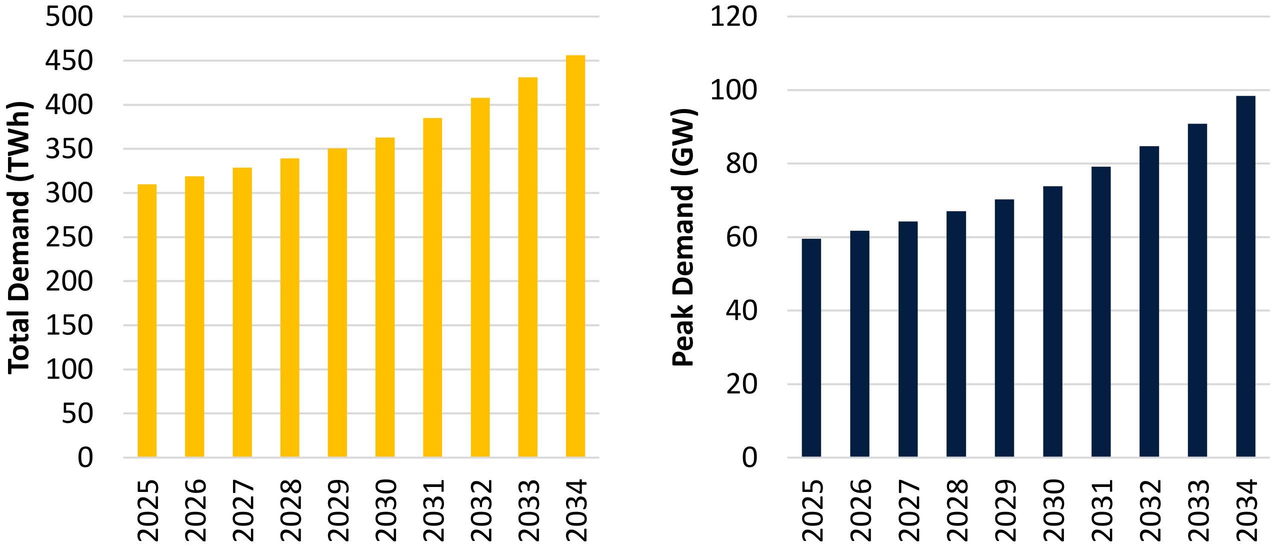
Annex O of the Energy and Emissions Projections[24]. This scenario projects a significant increase in intermittent renewable capacity (wind & solar) over the period to 2034 as shown below.

Figure 10: *GB Capacity (GW) for 2025-34 from DESNZ Net Zero Higher Demand Scenario*



Total demand and peak demand for GB are also provided by DESNZ. This shows moderate increases in demand from 2025 to 2034 as shown below.

Figure 11: *Assumed GB total and peak demand for 2025-34 in the DESNZ Net Zero Higher Demand Scenario*



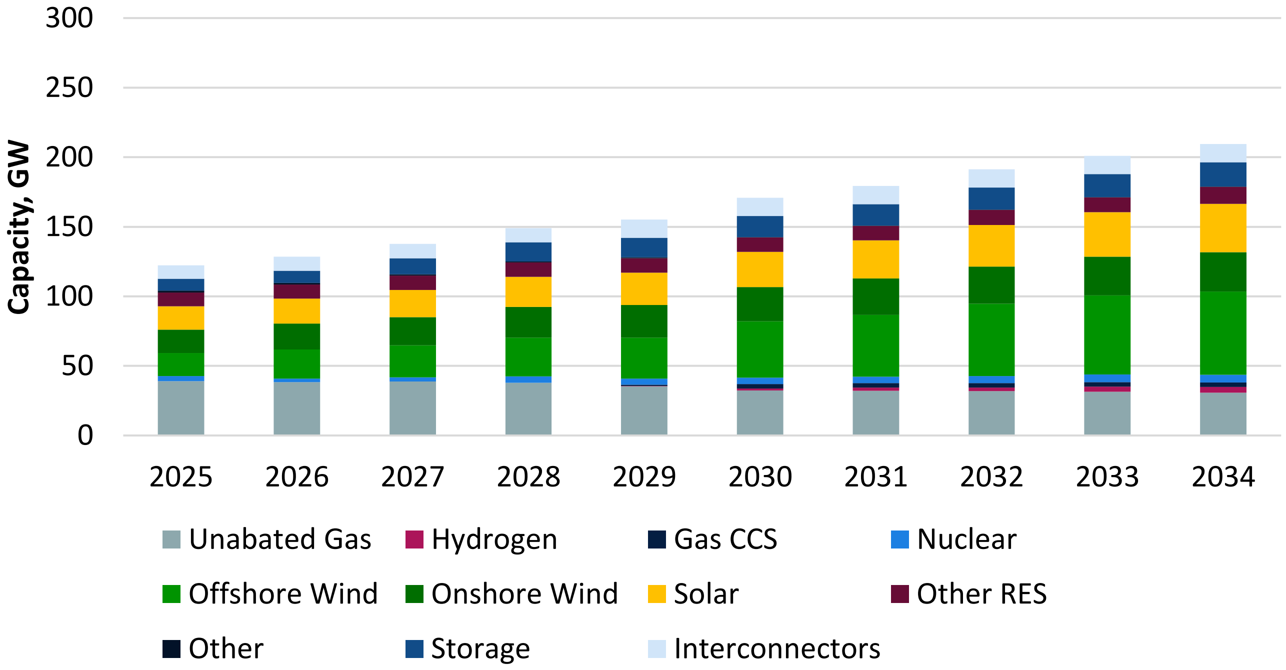
---

[24] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1059067/annex-o-net_zero-and-the-power-sector-scenarios.pdf

The assumptions for the FES 2022 System Transformation (ST) scenario were taken from the FES published documents[25]. The System Transformation scenario is one of four scenarios is the FES framework – it meets Net Zero by 2050 and reflects more change in the way in which we generate and supply energy rather than changing the way it is used. The FES was chosen as an alternative due to its ubiquity and due to its use in the Capacity Market target setting process, as well as to give a bigger contrast in possible outcomes compared to other DESNZ scenarios.

Compared to the DESNZ NZH scenario, the FES ST scenario has lower levels of demand and peak demand. In turn this translates into lower levels of total capacity on the system although it does also have high levels of renewable penetration in order to meet decarbonisation targets. The FES scenario also has a higher penetration of flexibility technologies, such as storage, Vehicle to Grid and Demand Side Response (DSR) which means demand is flattened and shifted more in the FES scenario compared to the DESNZ scenario. A lower heating demand will also contribute to flatter demand profiles.

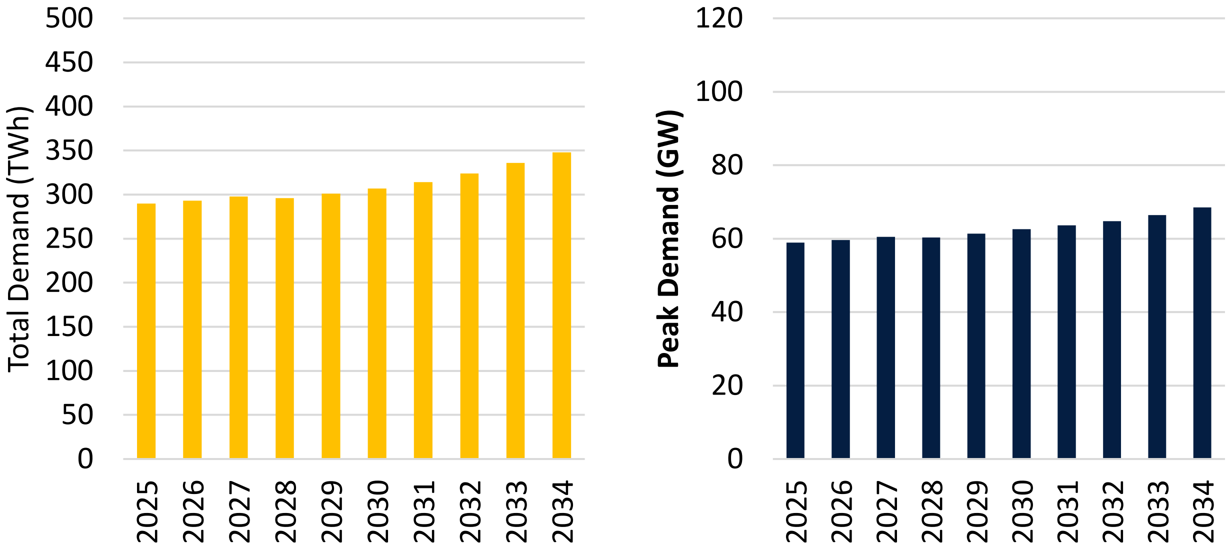
Figure 12: *GB Capacity (GW) for 2025-34 from FES 2022 System Transformation Scenario*



---

[25] https://www.nationalgrideso.com/document/263951/download

Figure 13: *Assumed GB total and peak demand for 2025-34 in the FES 2022 System Transformation Scenario*



It should be noted that not all details of the FES scenario are publicly available. In particular, shifting the peak demand caused by electric vehicle charging and heating play an important role in these scenarios in maintaining security of supply, but NG ESO do not publish the resulting daily demand profiles. The published data does include some indication on the direction and magnitude of the shift, and these were incorporated in the daily demand profiles that were entered into the model.

Both scenarios are designed to meet security of supply to a given reliability standard. The DESNZ NZH scenario assumes that many small, low-load factor unabated gas plants are used to meet 3 hours LOLE, while the FES ST scenario relies more heavily on demand side response (DSR). They do not incorporate information on the likelihood of delivery of this capacity, as this is not their primary purpose – they are illustrative pathways that can be used to explore a particular energy system in the future and how it responds to changes. The scenarios do not precisely meet the LOLE standard of 3 hours/year – in the case of DESNZ NZH this is due to the lumpiness of new build and the modelling of a demand curve in the Capacity Market. To ensure scenarios and years were compared on a consistent basis in this modelling, small amounts of firm capacity was added (or subtracted) to ensure precise alignment in each year.

A typical system stress event looks quite different over time across the two scenarios, as illustrated in the graphs below (Figure *14* to Figure *17*). In the 2025 DESNZ NZH example, generation over the evening peak is not enough to meet demand and this peak is the cause of the event. A large drop in total available generation is visible in the event in 2034, this is a drop in wind generation during the day. However, the evening peak remains the cause of the event, as there is still enough capacity available to meet the demand during the day. It is worth noting that total capacity on the system increases: peak demand in 2025 is less than 60GW, but it is almost 110GW in 2034.

Figure 14: *An example system stress event in the DESNZ Net Zero Higher demand scenario in 2025, modelled with current reliability standard (LOLE 3 hours/year)*

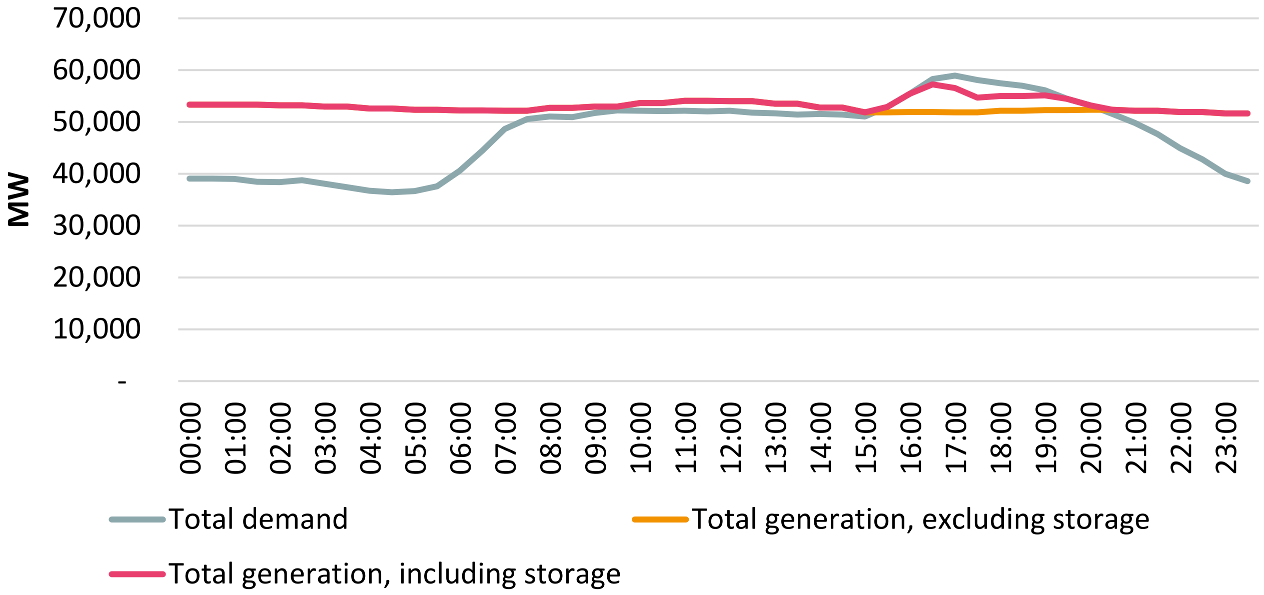
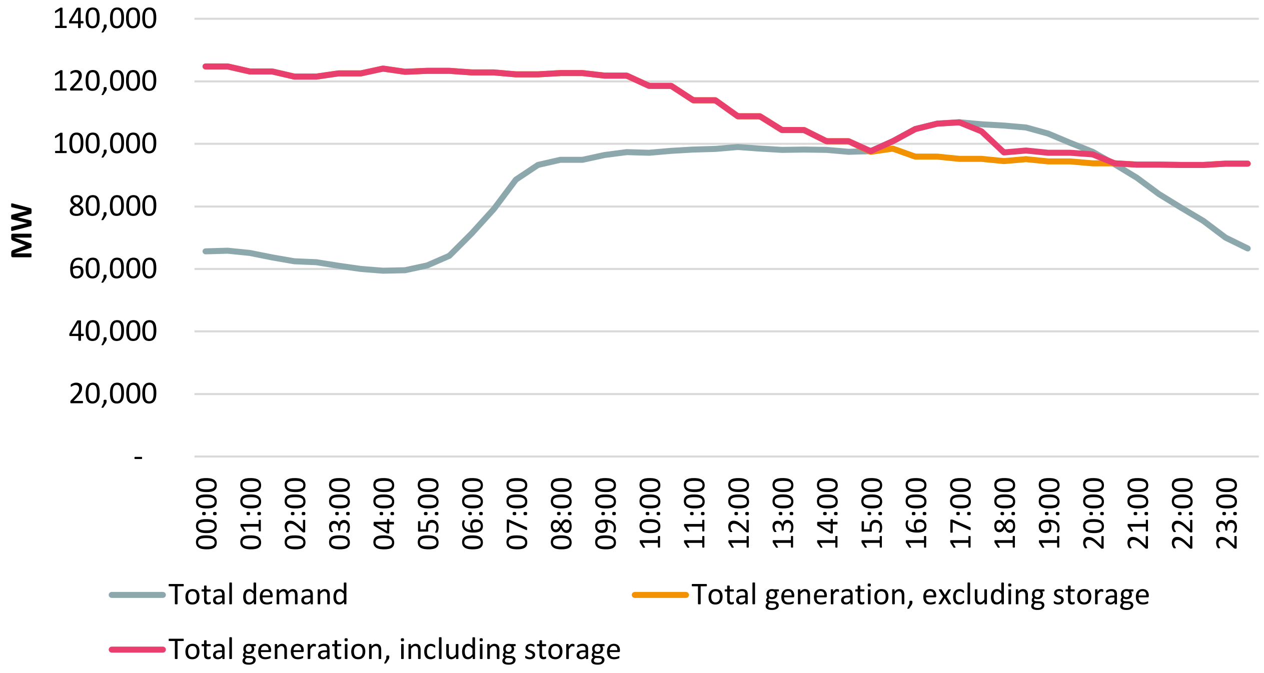


Figure 15: *An example system stress event in the DESNZ Net Zero Higher demand scenario in 2034, modelled with current reliability standard (LOLE 3 hours/year)*



The situation is often different in the FES System transformation scenario. In the example events of both 2025 and 2034, generation without storage is lower than demand for long periods over the middle of the day. However, large volumes of storage relative to demand in the FES scenario means storage plays a greater role in topping up the generation over this period, before running out to meet the evening demand peak. Demand is also closer to generation for larger parts of the day in this scenario, particularly in 2034. This means that a reduction in generation due to lower renewables output is more likely to cause a system stress event. At the same time however, the typical system stress event is still occurs during the evening peak period with the highest demand.

Figure 16: *An example system stress event in the FES System transformation scenario in 2025, modelled with current reliability standard (LOLE 3 hours/year)*

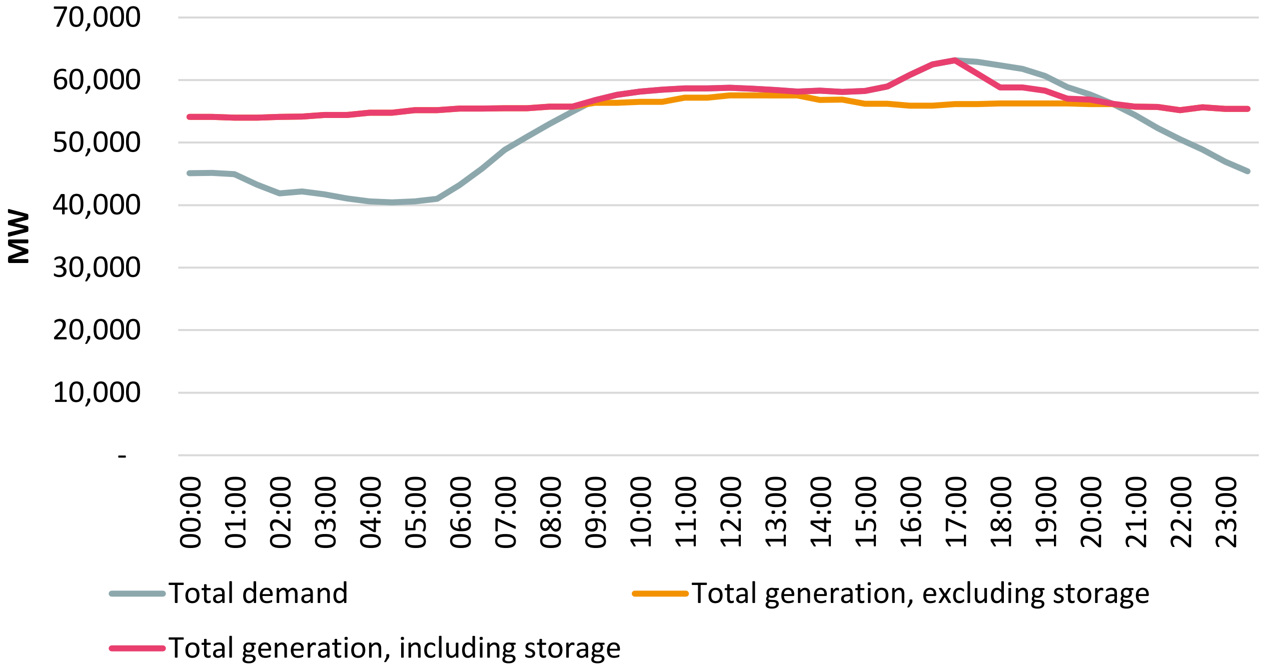
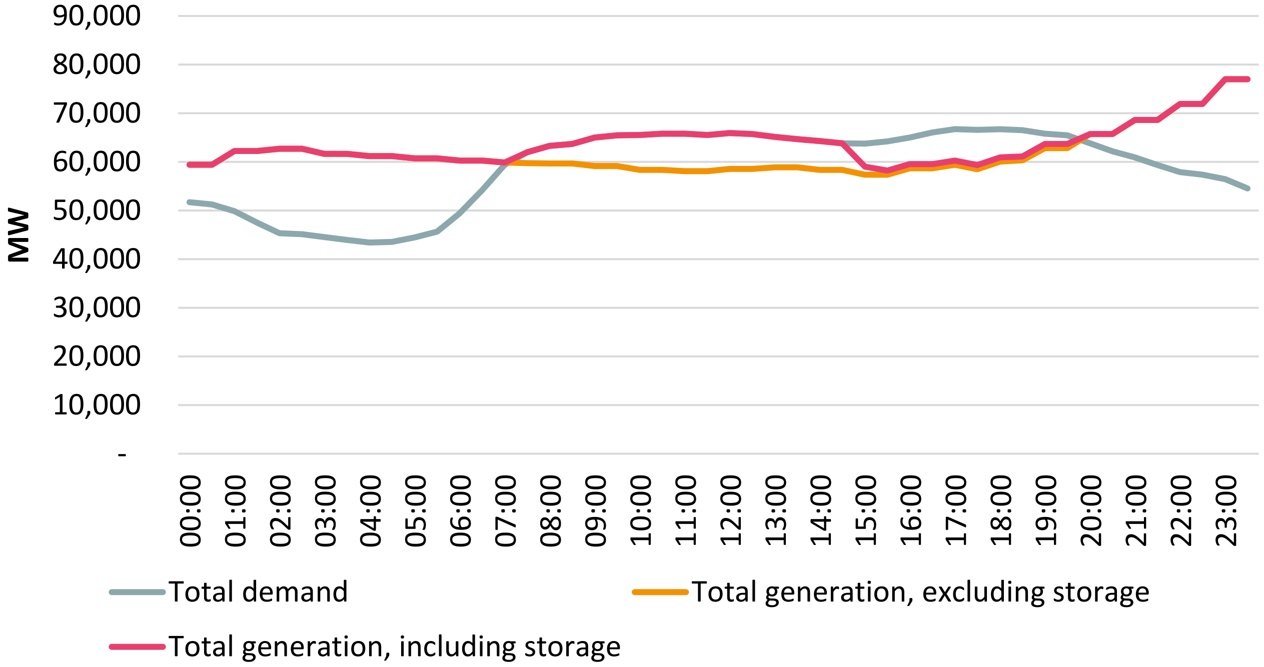


Figure 17: *An example system stress event in the FES System transformation scenario in 2034, modelled with current reliability standard (LOLE 3 hours/year)*



4. Modelling Results

This chapter outlines the results from the modelling undertaken, showing how key metrics change over time and the impact of using alternative reliability standards.

4.1. Will security of supply metrics change in future?

For the first part of the modelling, we assess how each of the key metrics identified through the literature review change from 2025 to 2034 under each scenario. In order to enable closer scrutiny of how the other reliability metrics are changing over time as the system changes, the average total length of system stress events (LOLE) in each year has been fixed to 3 hours/year. This is in line with the existing GB reliability standard that informs target capacity levels for the Capacity Market.

Loss of Load Expectation (LOLE)

While the LOLE has been fixed, it is possible that the underlying distribution it is based on, the total length of system stress events in a single year, changes over time. For example, the average total length of system events can remain 3 hours/year if the proportion of simulations with longer than 1 hour of events/year increases, and the proportion of 0 hours/year goes down.

The DESNZ NZH follows this trend, with a small decrease over time in the proportion of simulations with no event at all. There are small increases in the total events that are 1 hour or longer in duration. Overall, though, these changes are relatively small. There is more variation over time in the FES ST scenario with the proportion of simulations with no event at all increasing substantially from 2025 to 2034 but the proportion of simulations with total events of any length coming down slightly, with somewhat larger changes for events that are shorter. Overall, however, the changes are relatively small in the distribution of LOLE from 2025 to 2034, except for the proportion of simulations with no events at all.

Figure 18: *Total annual system stress event distribution (LOLE) over time, DESNZ Net Zero Higher Demand scenario*

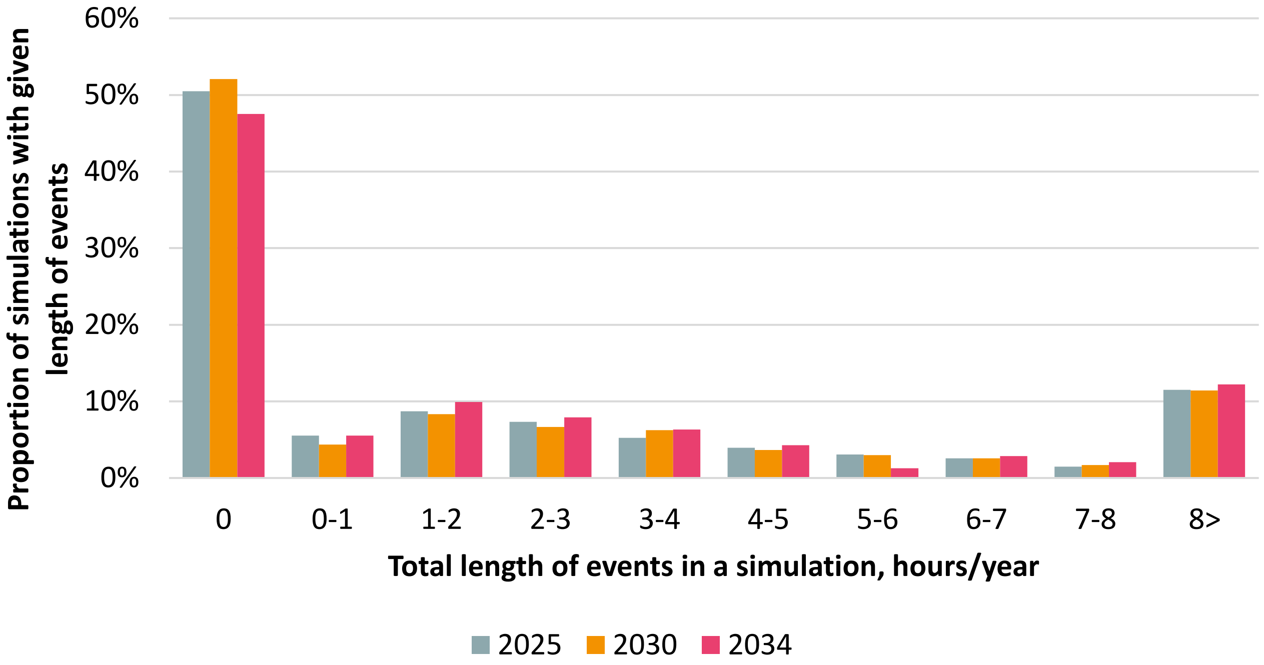
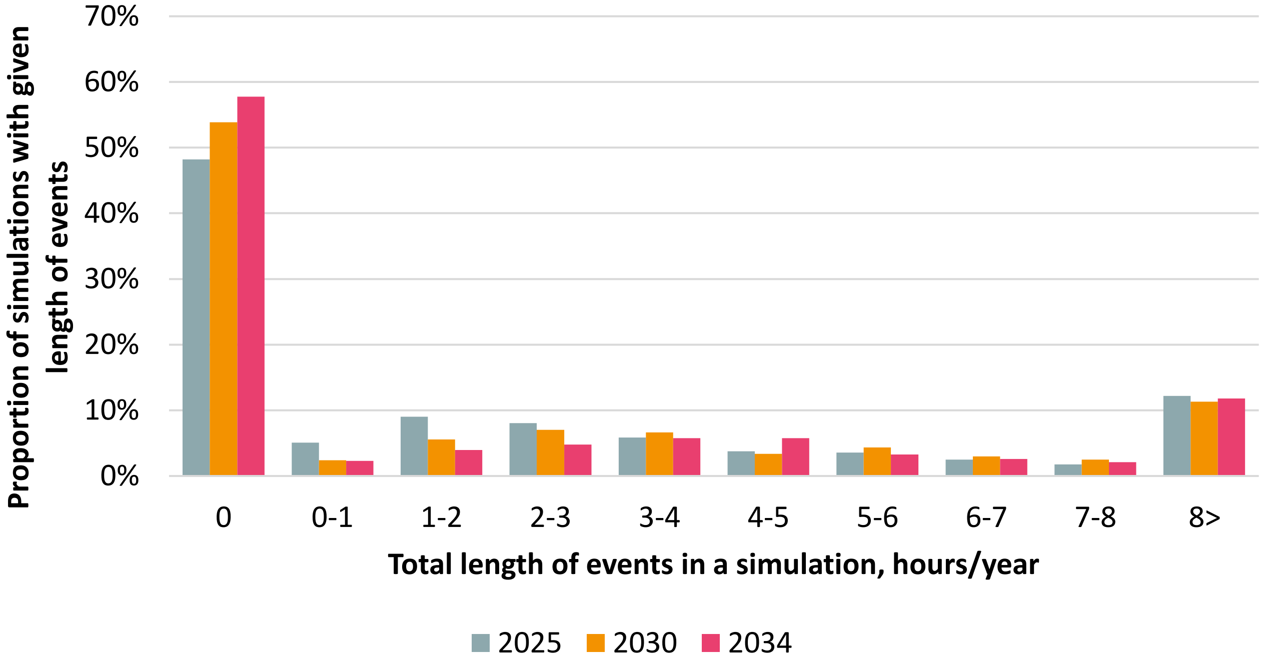


Figure 19: *Total annual system stress event (LOLE) distribution over time, FES System transformation scenario*

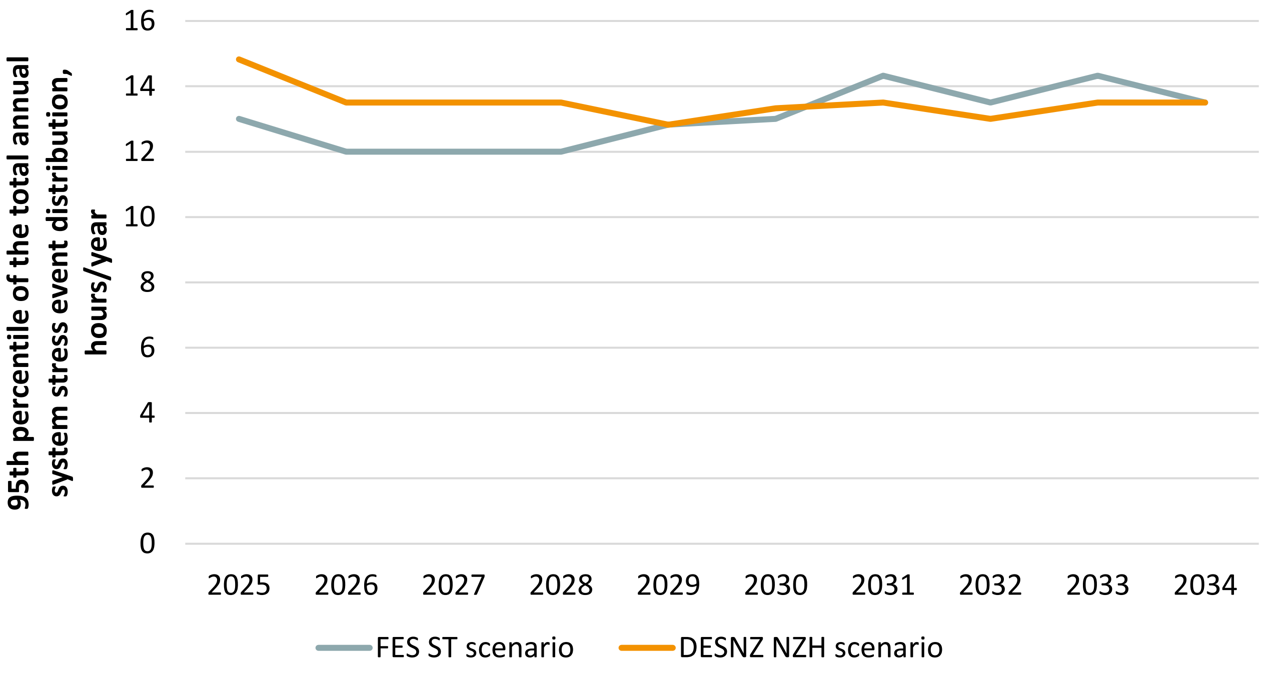


The distributions can be summarised using a variety of summary statistics, including the 95th percentile, which can be seen as a possible metric for volatility as outlined in 2.3. Conceptually, this value corresponds to the likelihood of an event occurring in 1 in 20 years (for example, 14 hours/year means that in 1 in 20 years, the total annual system stress events will be 14 hours or longer). This is an important metric characterising the distribution of events, but in the two tested scenarios, it changes relatively little over time. This suggests that extreme years, with more total hours of simulated system stress events, do not become more likely over time (when LOLE is held constant at 3 hours/year). This means that using LOLE95 as an alternative reliability standard metric would not result in significant differences to keeping LOLE as a metric in the scenarios explored. For this reason, we have not explored it further as a second reliability standard metric in combination with LOLE.

The observed lack of change in the probability of extreme outcomes can potentially be explained by a few factors.

- First, the most extreme simulations will likely be based on the same underlying low-wind and high-demand sample years, as there is a limited number of sample years used in the modelling. These are likely to drive extreme simulations with a similar number of events in future years.
- Secondly, there is a fundamental link between LOLE and LOLE95. The most extreme events contribute a disproportionate amount to LOLE, with around 50% of simulations having no events, and the top 5% of simulations contributing to about 1/3rd of LOLE. Therefore, any significant increase in LOLE95 would flow through into LOLE, so the assumption that LOLE is held constant (and firm capacity is procured to ensure this) will help mitigate any increases in LOLE95.
- Finally, we do not assume any future changes in the underlying variability of weather (e.g. wind patterns), due to climate change or other factors. Capturing these changes could impact the probability of extreme events occurring in the future . Addressing this would need further work on how variability of these input factors could change in the future.

Figure 20: *95th percentile of total annual system stress event hours (LOLE) over time and between scenarios*



As discussed earlier in this report, two parameters of a system stress event scenario, the frequency of events in a year and the duration of an event, directly explain the total annual length of events in a year. The expected value of each of these distributions are LOLF, LOLD and LOLE respectively, and the relationship between these hold in this context, as well: LOLE = LOLD * LOLF. Therefore, as long as the total annual length of events is fixed (LOLE) as it is in the existing reliability standard, the frequency of events (LOLF) and the duration of events (LOLD) must balance each other out – if the frequency of events goes up, the duration of a single event must go down, and vice versa. How these metrics change over time is explored below.

Loss of Load Frequency (LOLF)

The LOLF represents the expected number of continuous events (frequency) in a year where there is insufficient supply available in the market.

The LOLF distribution changes relatively little in the DESNZ NZH scenario: there is a drop in the probability of a year with no events at all, and a corresponding increase in the probability of a year with exactly one event. The probability of a year with more than 1 event does not change substantially over time. The summary statistics of the distributions also do not change substantially over time, with a small decline in LOLF and the 95th percentile over time and no large change in the median.

Figure 21: *System event frequency distribution (LOLF) over time, DESNZ NZH scenario*

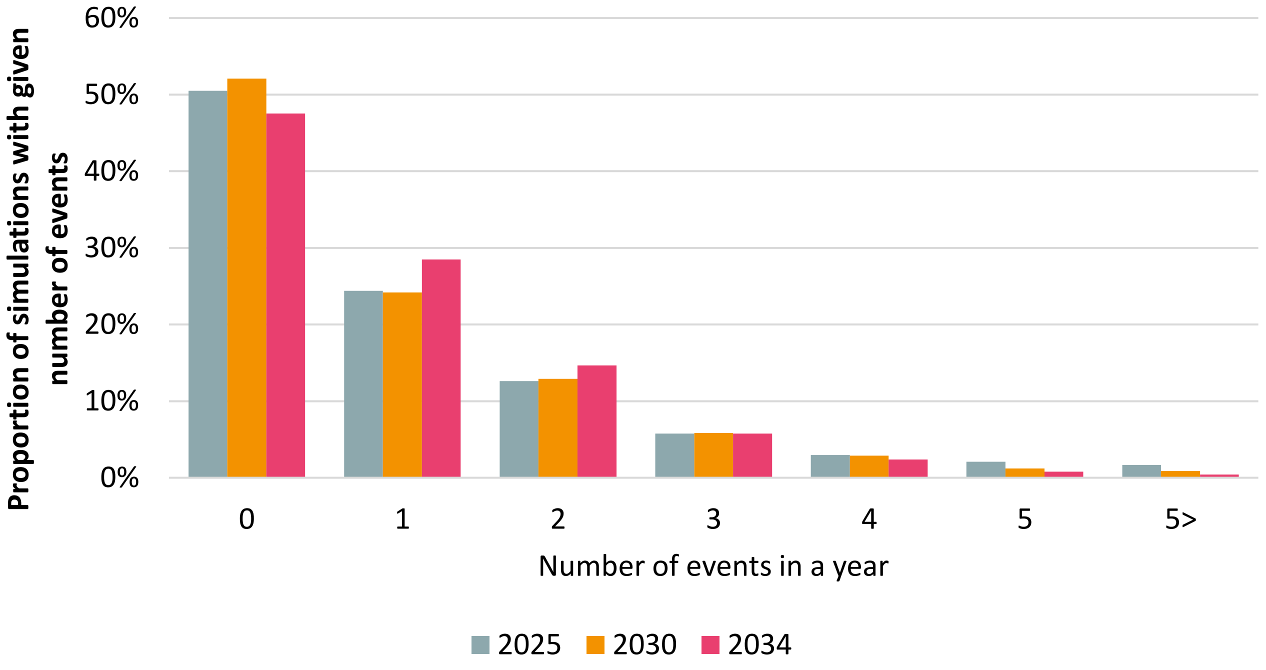
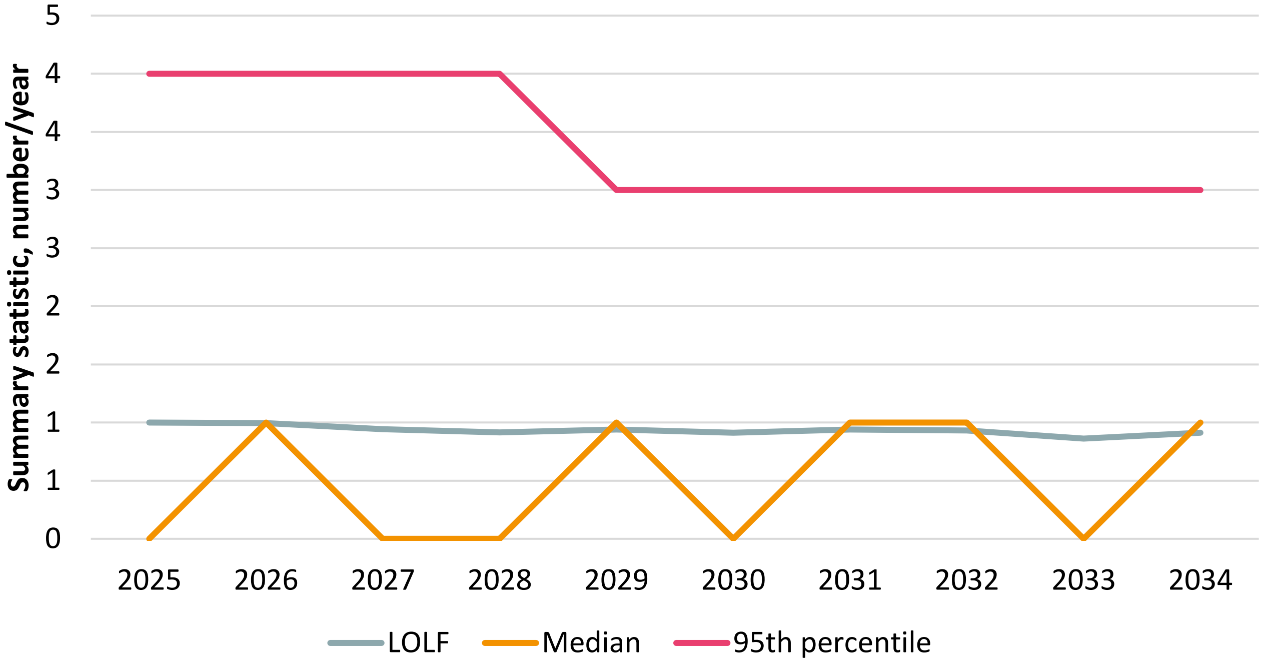


***Figure 22**: Summary statistics of system event lengths over time, DESNZ NZH scenario*



A different behaviour is observed FES ST scenario where the proportion of simulations with no event at all steadily declines. The proportion of simulations with exactly one event stays relatively stable while the proportion of simulations with two or more events decreases. As the LOLE distribution is maintained in the modelling as it would be in reality via the capacity mechanism, this behaviour balances out the changes observed in the LOLD distribution which are caused by a change in the generation capacity mix and the daily demand profile (described further below). The summary statistics of the distribution also show some changes over time, with the LOLF and the 95th percentile of the distribution both clearly trending downwards.

***Figure 23:** System event frequency distribution (LOLF) over time, FES ST scenario*

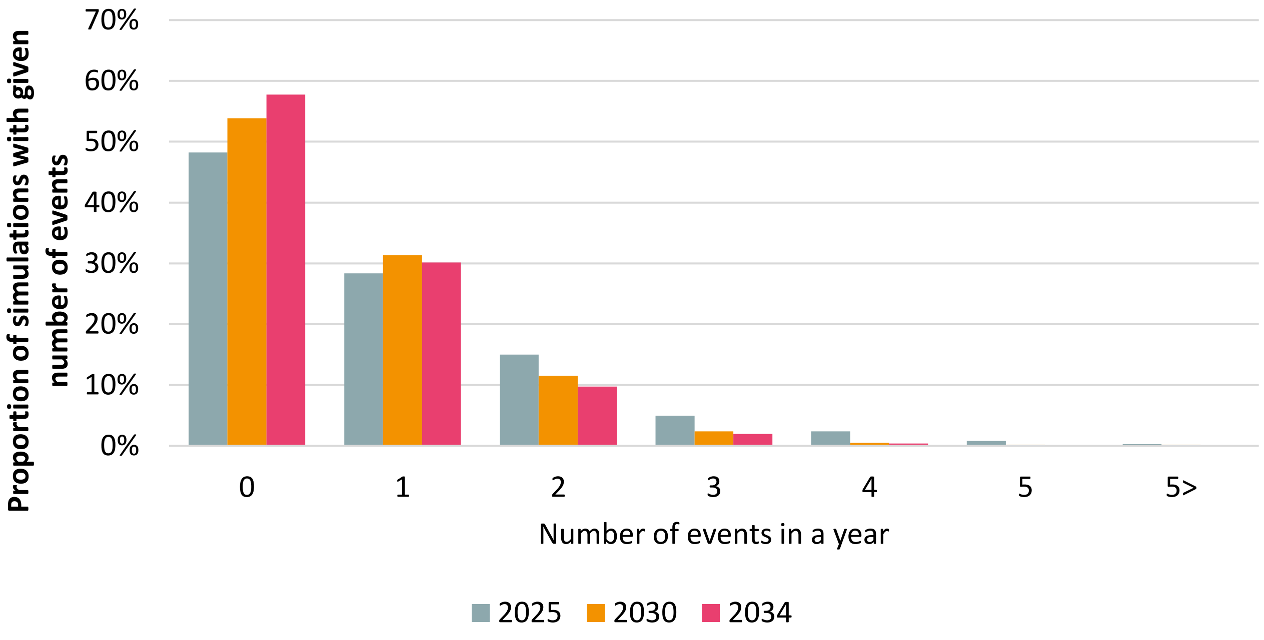
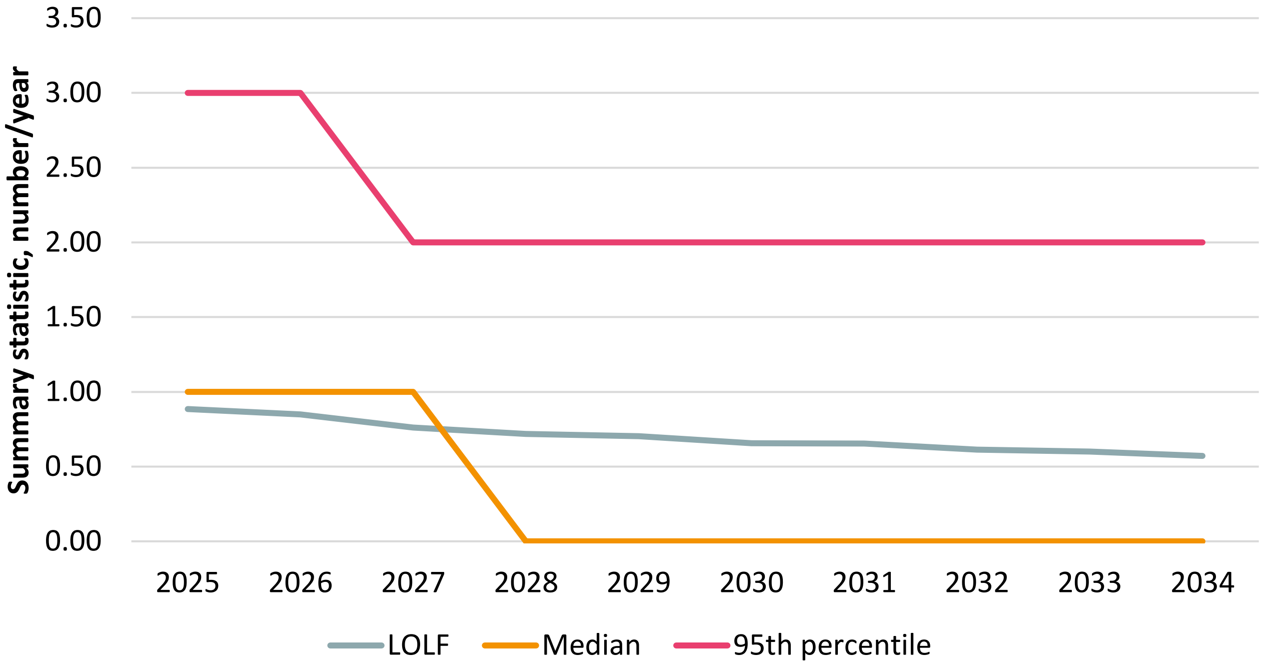
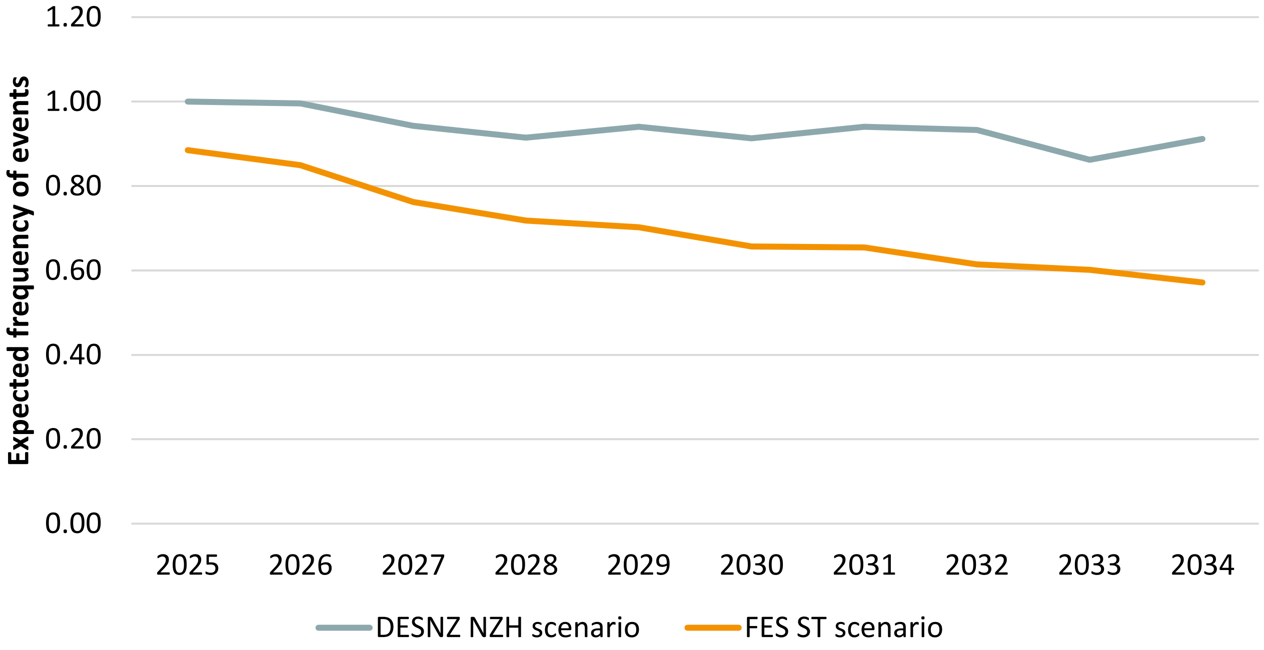


***Figure 24**: Summary statistics of system event lengths over time, FES ST scenario*



The changes in distribution of LOLF across the two scenarios over time is also reflected in the average number of events in a single year, as shown below. LOLF is slightly declining across all years in the DESNZ NZH scenario but the decrease is much more evident in the FES ST scenario from an average of 0.88 events in 2025 to 0.57 events in 2034. Given that the LOLF decreases over the years in these scenarios, this metric has not been taken forward as a potential alternative metric in this study.

***Figure 25**: Average number of events in a single year (LOLF) over time and between scenarios*



Loss of Load Duration (LOLD)

The LOLD represents the average length of an event in a year where there is insufficient supply available in the market. Compared to the previous two metrics (LOLE and LOLF), the distribution of LOLD shows some increase over time, especially in the FES ST scenario.

In the DESNZ NZH scenario, most events are less than 3 hours in length, although over 10% of all events are longer than 6 hours across all simulations. This distribution changes over time with a drop in events 2-3 hours in length and a sharp increase in the proportion of events that are 6 hours or longer from 10% to 20%.

***Figure 26**: Distribution of different system event lengths (LOLD) over time, DESNZ NZH scenario*

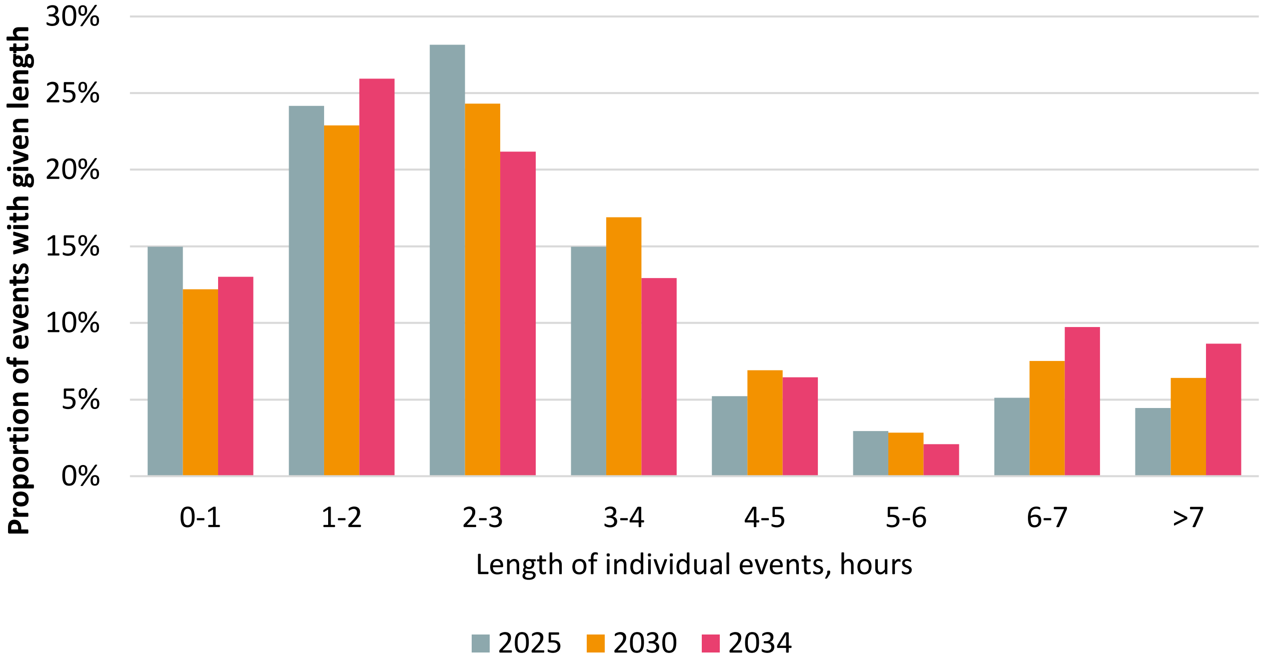
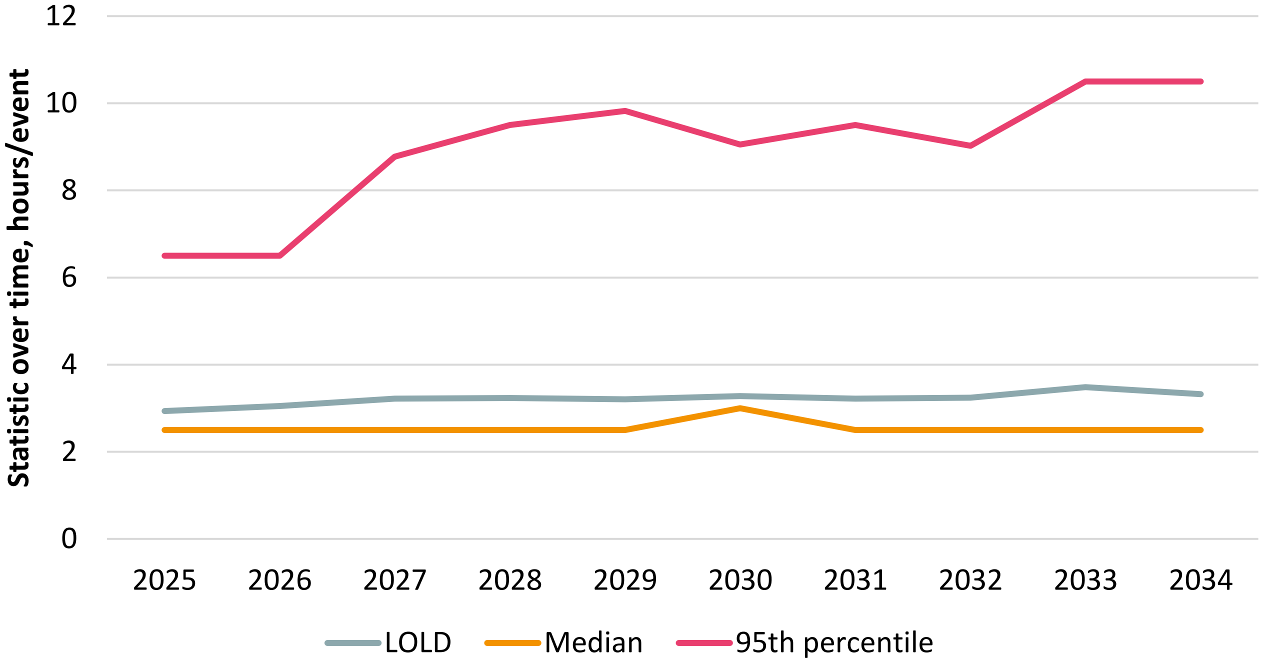


Figure 27: *Summary statistics of system event lengths over time, DESNZ NZH scenario*



There is more variation in the FES ST scenario, where there is a higher probability of longer events – even in 2025, the probability of an event lasting more than 6 hours is over 10% and it rises to over 33% by 2034. Interestingly, the average event length increases substantially over time, but the 95th percentile of the distribution, which captures extreme values, changes less. This corresponds to the changes observed in LOLF where the distribution shows a higher proportion of simulation with less than 1 event.

Figure 28*: Distribution of different system event lengths (LOLD) over time, FES ST Scenario*

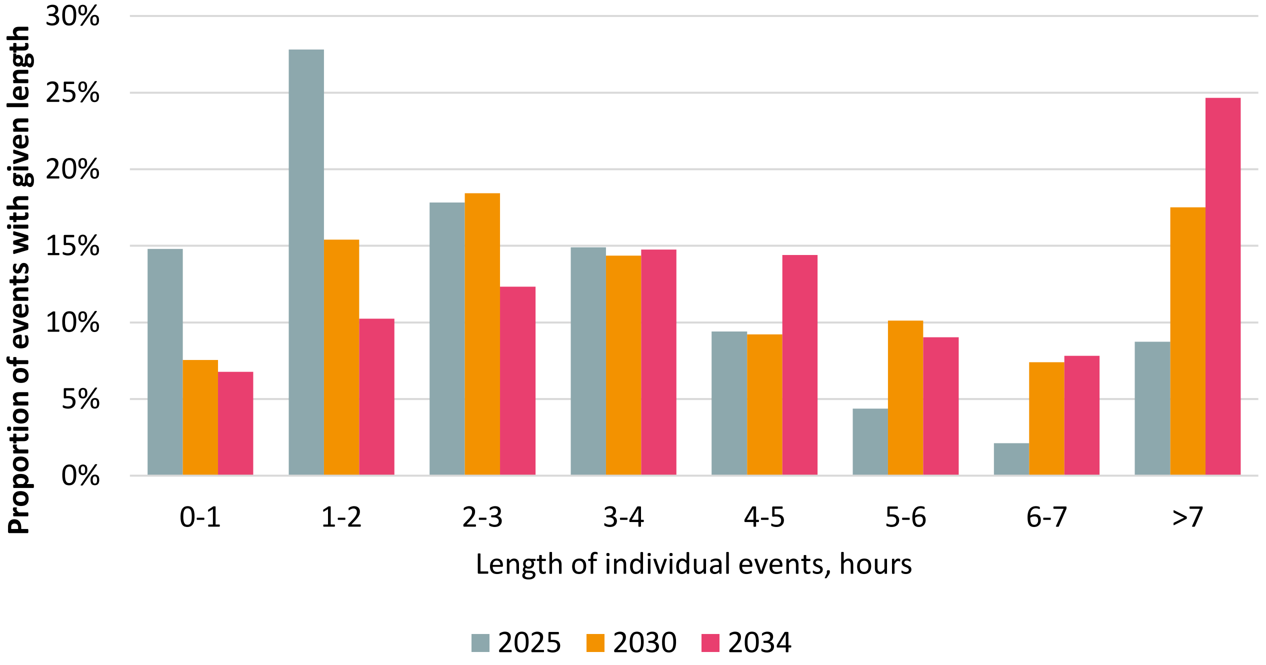
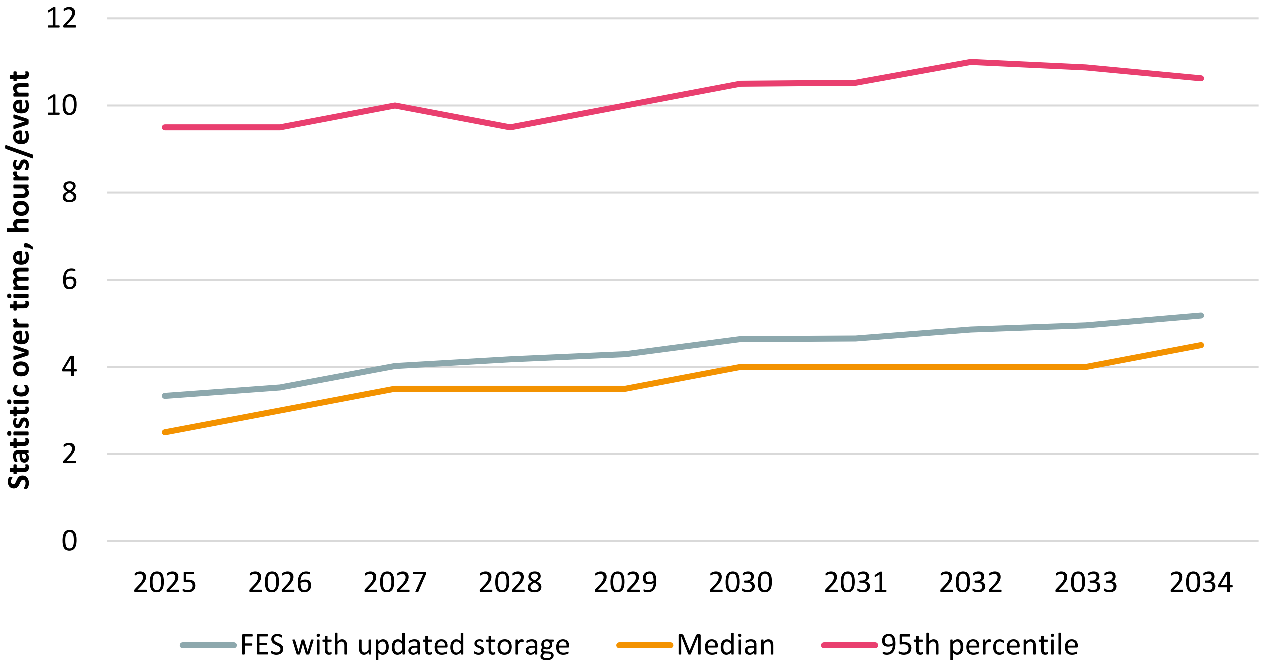
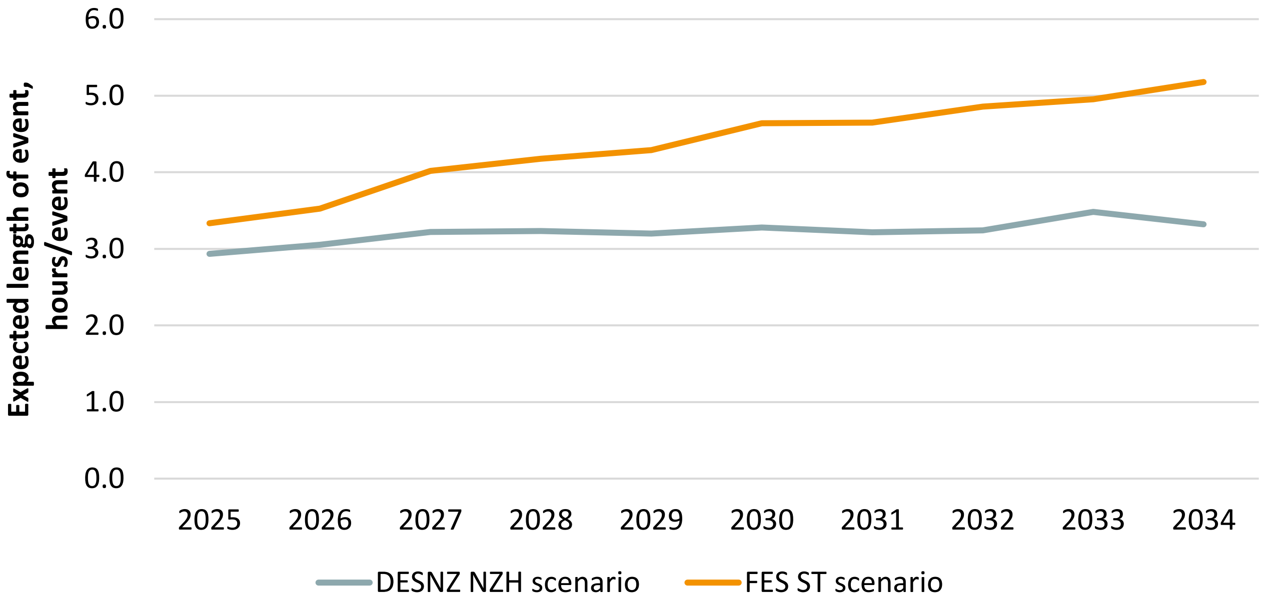


Figure 29: *Summary statistics of system event lengths over time, FES ST scenario*



The difference between the two scenarios is likely related to the flatter daily demand profile and the higher volume of storage capacity in the FES ST pathway. The former means that on days with lower total available generation due to outages and low wind and solar generation, the flatter daily demand peak can cause longer system stress events. The latter means that more storage capacity is available to cover shorter events. Together, these leads to low intermittent renewable generation playing a larger role in driving system stress events in future years. The average length of an event reflects this behaviour, as the value is increasing at a much faster pace over time in the FES ST scenario, than in the DESNZ NZH scenario. This means that in the former scenario, when a system stress event does occur, it will be typically longer.

Figure 30: *Average length of an event (LOLD) over time and between scenarios*



The increasing average length of an event (LOLD) means that in a future energy system, consumers may experience different system events than what has traditionally been assumed. As the LOLD is increasing in the FES ST scenario, for this scenario it is worth exploring this metric as an alternative reliability standard. This is explored further in section 4.2.

Expected Energy Unserved (EEU)

Consumers use and pay for each MWh unit of energy in practice, therefore the amount of energy not served is an important metric. The EEU represents the amount of electricity demand that is expected to not be met by available supply over the course of the year. Across both scenarios modelled, there is a clear increase over time in the average energy unserved in a single year, as well as in the 95th percentile of the metric. The rise in the DESNZ NZH scenario is higher than the rise in the FES ST scenario.

***Figure 31**: Summary statistics of unserved energy over time, DESNZ NZH scenario*

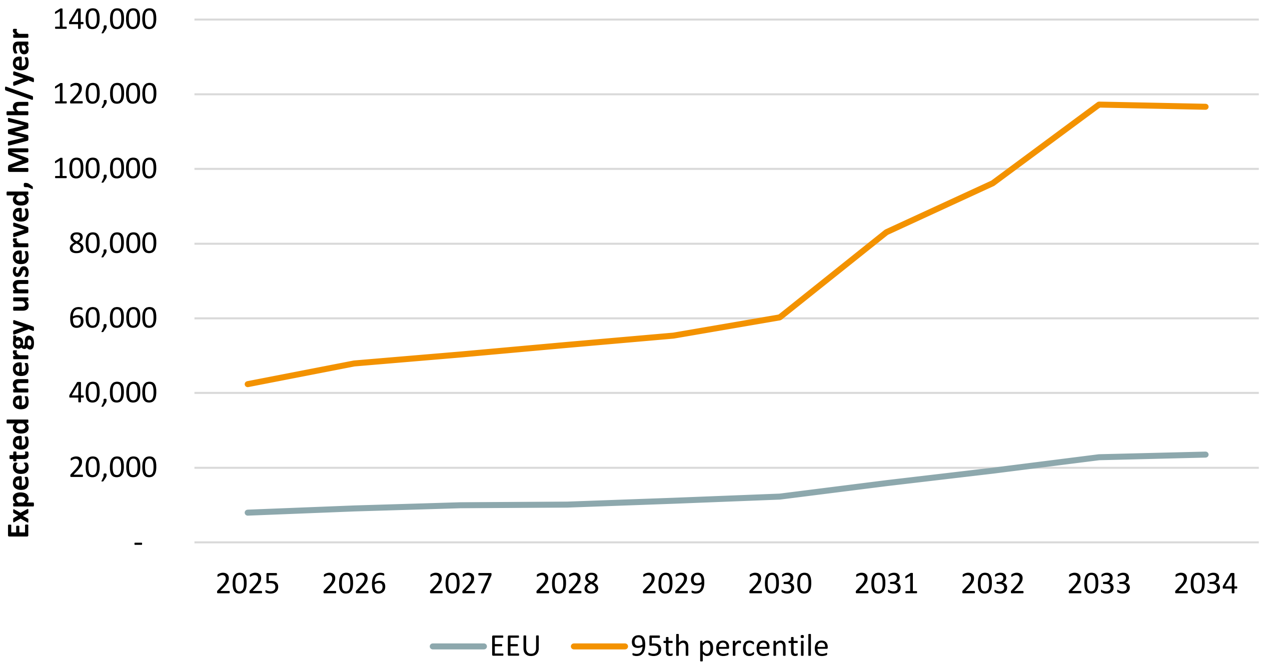


Figure 32*: Summary statistics of unserved energy over time, FES ST scenario*

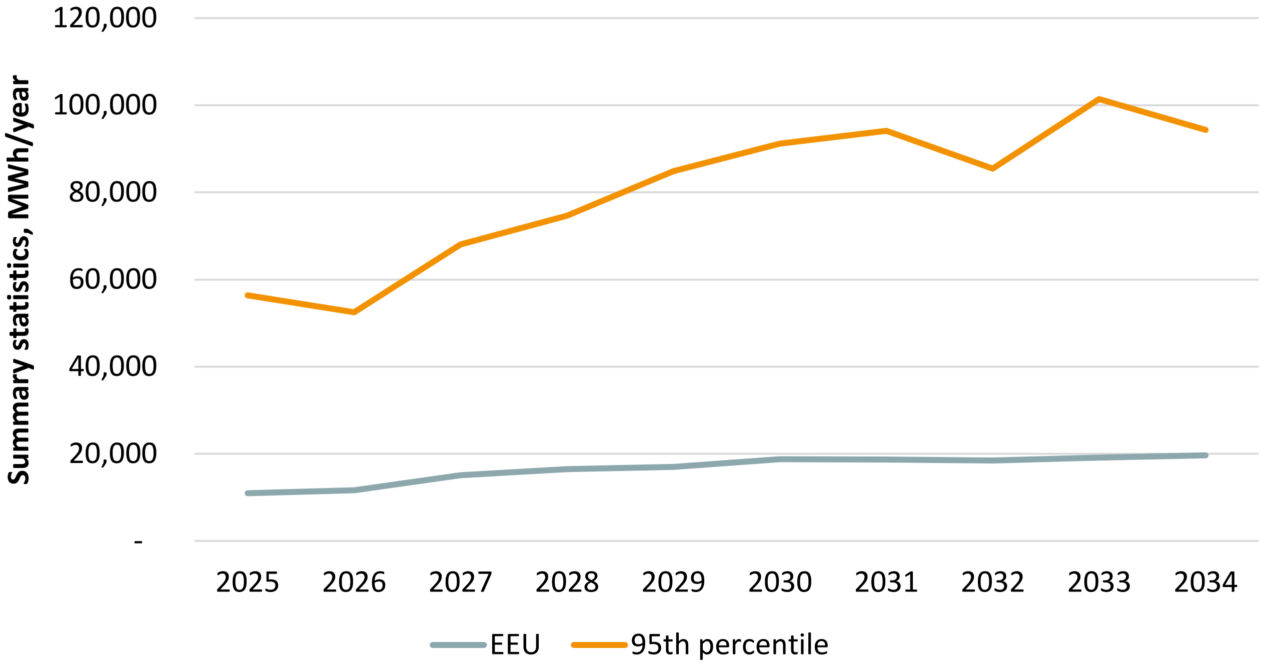
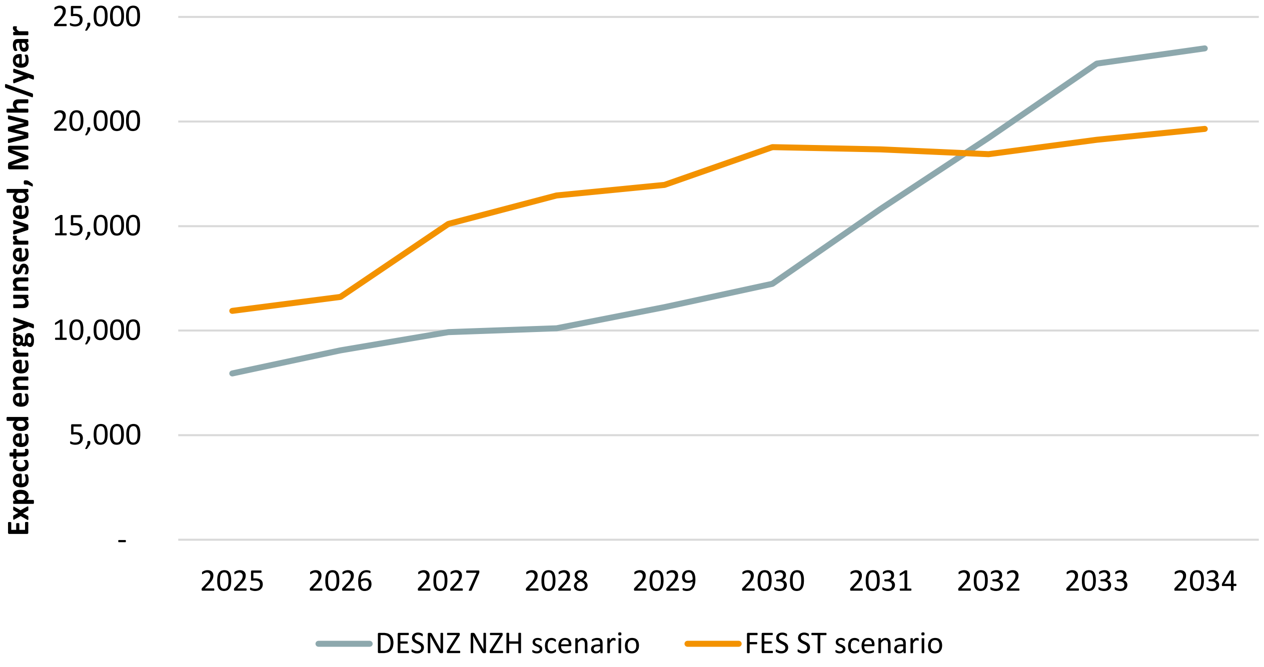


Figure 33: *Expected energy unserved (EEU) in a year across both scenarios*



Peak demand and total demand are both increasing over the modelled time period as sectors such as transport and heating are electrified. As the power system becomes larger and more energy is generated and used, we would expect to see numerically larger swings. Normalising the EEU in each scenario and year with the relevant peak demand tests if there is any variation in the depth of events relative to demand. This shows a slight increase in the normalised EEU over time and a considerable increase in the 95th percentile in the DESNZ NZH scenario, with a similar increase in the FES ST scenario. This mirrors the relative increase observed in the

95th percentile of the length of system events distribution, as the rare longer events are going to influence this distribution, as well. Both link back to the increase in storage capacity volume, as this can cover shorter and shallower events, leaving only longer and deeper events.

Figure 34: *Summary statistics of normalised unserved energy over time, DESNZ NZH scenario*

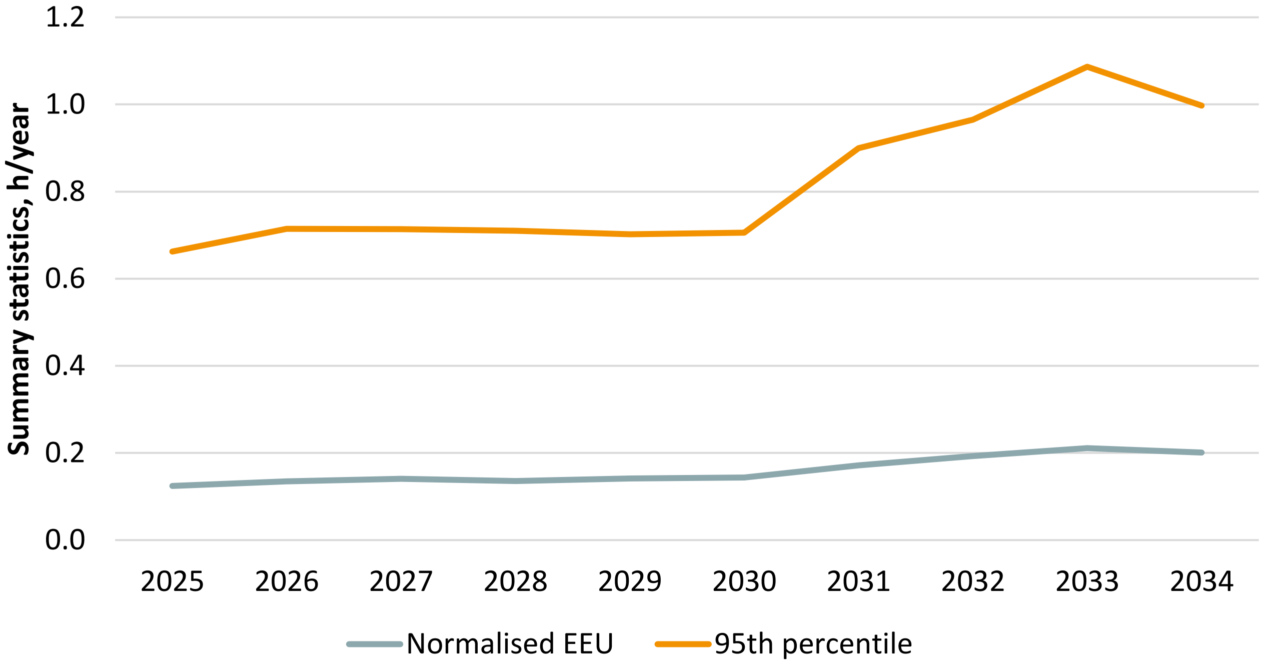


Figure 35*: Summary statistics of normalised unserved energy over time, FES ST scenario*

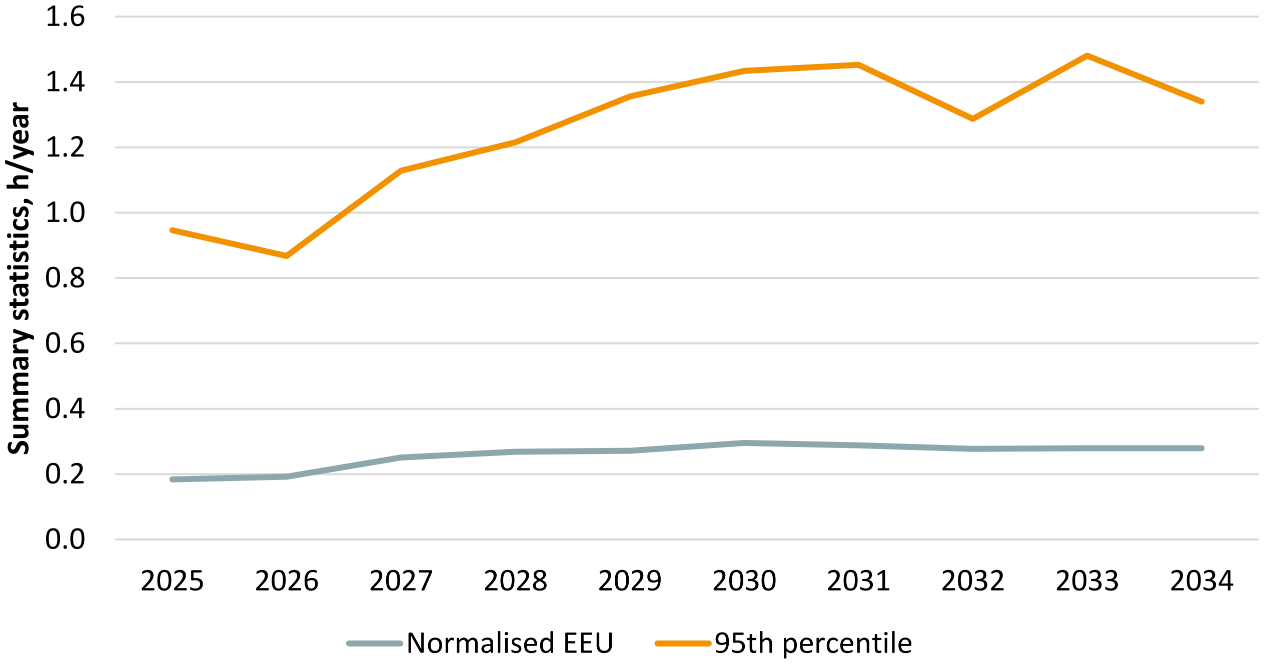
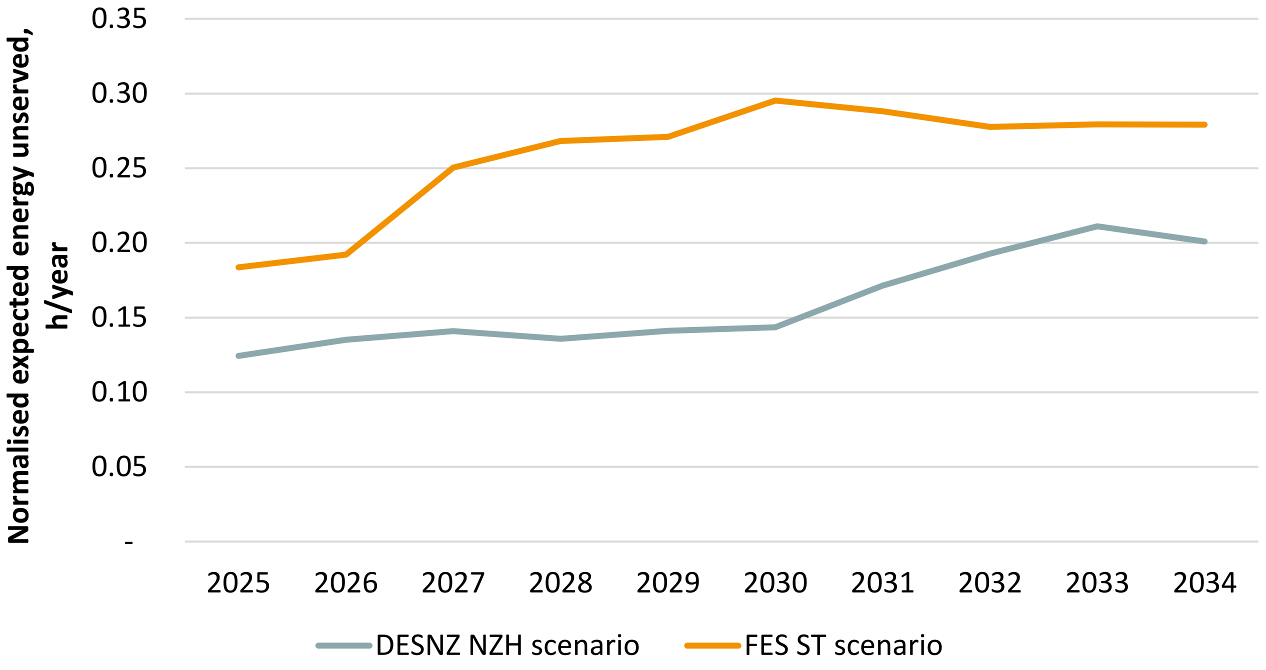


Figure 36: *Normalised expected energy unserved (EEU) in a year over time and between scenarios*



Given EEU increases over time while LOLE is held constant, the average depth of an individual system stress event will increase over time, in particular in the DESNZ NZH scenario. The Electricity System Operator (ESO) has traditionally been able to address certain depths of events without needing to resort to demand disconnections, but larger losses may require new products and approaches (for example, network reinforcements).

This additional unserved energy is distributed over a larger overall demand. Instead of a fundamentally new experience for consumers in terms of the depth of events, system stress events will affect a greater breadth of electricity demand, particularly across transport and heating. This is shown through the finding that on a normalised basis, EEU changes much less over time. As the increase in EEU is in large part due to the increase in peak demand, it may not be a good candidate for an alternative reliability standard in the GB power system. Using a normalised EEU resolves this issue, by presenting EEU as a proportion of peak demand. The results show that normalised EEU does increase over time, therefore it may serve as a candidate for an alternative reliability standard. However, its interpretation is less intuitive than LOLD and it changes less over time, therefore it was not chosen in this study. The behaviour of EEU was, however, closely observed when using LOLD as a combined metric with LOLE, described further below.

Critically Tight Events

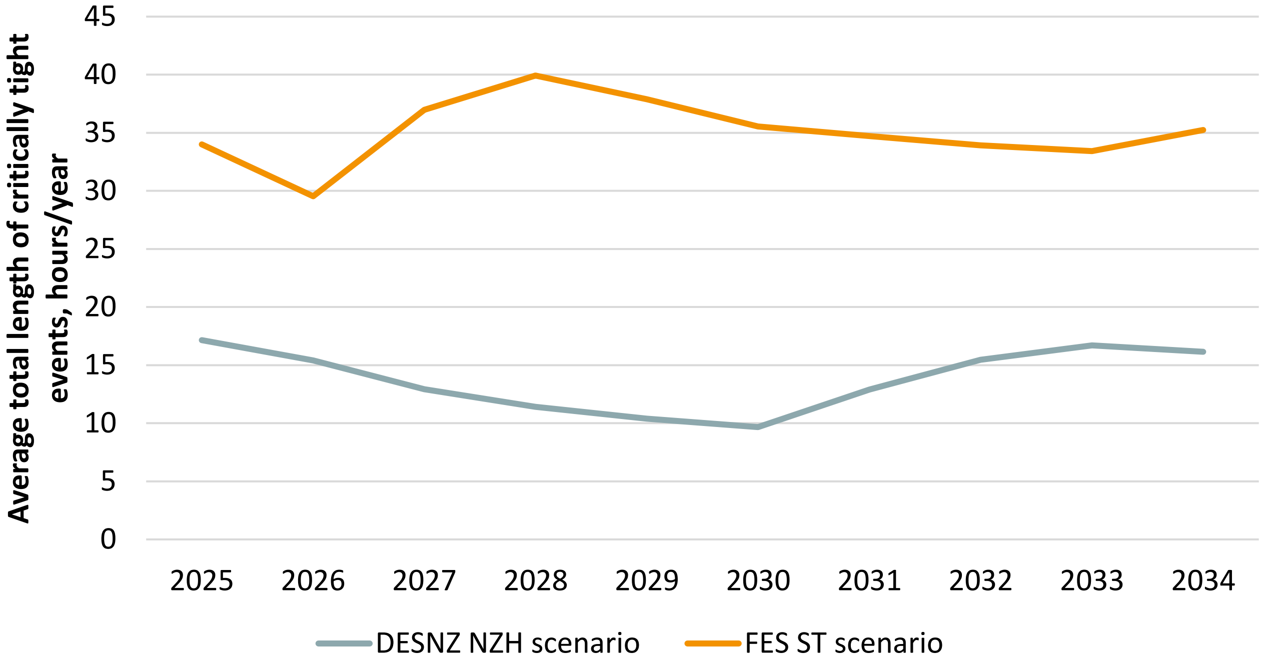
System stress events represent a critical situation where demand is greater than supply, but the power system can also become stretched when total demand approaches total available supply. Wholesale prices would rise well above short run marginal costs and the resulting scarcity pricing would represent a significant cost to consumers. Depending on configurations of the network, it is possible that some parts of GB would need to be partially disconnected even if there is enough generation capacity on the system overall. As identified in the literature review, to look at this we have modelled how Critically Tight Events change over time. In this study, this is defined as an event for which demand is over 99% of the total available supply.

The average total length of critically tight events does not change substantially in the decade between 2025 and 2034 in either scenario but there is a difference between the two. This is likely due to the way that the UEM uses storage plants. In the UEM, storage plants only discharge during system stress events, reducing the length and frequency of the system stress events, they do not discharge during critically tight periods to reduce the tightness. In reality, storage plants would aim to capture the higher prices in the market and therefore reduce the overall tightness (notwithstanding days with both a critically tight period and a system stress event where limited duration storage plants would be incentivised to only discharge during the system stress event).

This means that the actual total length of critically tight events would likely be lower in reality as storage plants discharge, with the relative difference between total demand and total available storage capacity determining the real value. Modelling storage behaviour in this level of detail would require additional information on the system operator practices in critically tight periods, as well as information on the commercial practices of storage operators (dependent on the duration of storage, as well) – much of this is currently undergoing changes and will also be impacted by broader market changes. If this were fully incorporated into the modelling, it is possible that the initial difference observed between the two scenarios would be much less pronounced.

In addition, a critical factor other than 99% may return different results, depending on how the volatility of generation and demand change over time – flatter daily demand may lead to more critically tight periods if the critical factor is set lower and the generation mix remains the same.

Figure 37: *Average total length of critically tight events in a year over time and between scenarios*



Critically tight events are not system stress events, but instead a helpful metric on the general tightness of the system. These events do have a cost to consumers as the market prices are able to rise beyond the short run marginal cost, but setting the reliability standard of a capacity mechanism may have wider market impacts, as well. This, as well as the fact that the headline measure is unlikely to increase in either scenario tested suggests that this measure is not a good candidate for an alternative reliability standard metric.

Conclusions

From the two scenarios modelled, only some metrics show a noticeable change from 2025 to 2034:

- LOLE is assumed to be the reliability standard so is fixed at 3 hours. Across both scenarios, some minor changes are observed in the distribution but these are not consistent between the two. The 95th percentile of LOLE does not show significant change over time in either scenario.
- LOLD increases slightly in the DESNZ Net Zero Higher Demand scenario but the increase is much larger, from 3 hours in 2025 to over 5 hours in 2034 in the FES System Transformation scenario. This is a 55% increase in the average length of a system event compared to today, a considerable increase.
- LOLF shows a small drop in the DESNZ NZH scenario, but declines strongly in the FES ST scenario as events get longer on average and occur less frequently.
- EEU increases over time in both scenarios. However, once normalised for the increase in peak demand, EEU changes much less. This indicates that although the absolute

magnitude of the depth of the events increases, the depth of events relative to overall demand levels will not change substantially over time.

- Critically Tight Events hold steady over time in both the DESNZ NZH and FES ST scenarios. Critically Tight Events are higher in the FES ST scenario as it includes proportionately more storage capacity, as well as flatter demand, meaning that generation is more often close to demand.

Overall, more change is observed across the metrics in the FES ST scenario compared to the DESNZ NZH scenario. This is due to the nature of system stress events changing more in the FES ST scenario. The FES ST scenario has a flatter daily demand profile meaning that on days with lower total available generation due to outages and low wind and solar generation, the flatter daily demand peak can cause longer system stress events. This scenario also includes proportionately higher storage capacity, which can cover more of the shorter events, leaving only longer events to be registered as system events. This essentially means that it is low intermittent renewable availability driving system stress events in future years more so than peaks in demand.

Across many of the metrics modelled, either a decline or no significant change is observed in both scenarios. This means that these metrics are not good candidates for an alternative reliability standard as they either scale with LOLE or are already declining without any change to the reliability standard. The two exceptions to this is LOLD, especially in the FES ST scenario, and normalised EEU. The change is more consistent over the years for LOLD and its interpretation is more intuitive, therefore this metric is taken forward into the second part of the modelling where alternative reliability standards are modelled. However, we closely monitor the way in which the normalised EEU changes under the combined metric to see what insights can be gained from its behaviour.

4.2. Modelling alternative reliability standards

The literature review identified that a combination of metrics could be a useful alternative to the existing reliability standard as it enables more complex stress events to be targeted. Given this finding, the LOLD metric that changes in the modelling is combined with the existing LOLE metric as an alternative reliability standard to understand the impact this change could have. In this section, we compare the level of derated capacity needing to be available for the existing reliability standard and the alternative approach. We also look at the impact on the key security of supply metrics and consider how changes may affect other elements of the Capacity Market such as derating factors for certain technologies.

The two reliability standards modelled are:

- **Option 0: Loss of Load Expectation (LOLE)** – this option keeps the current reliability standard. This is modelled to enable a comparison of the alternative reliability standards to the status quo.

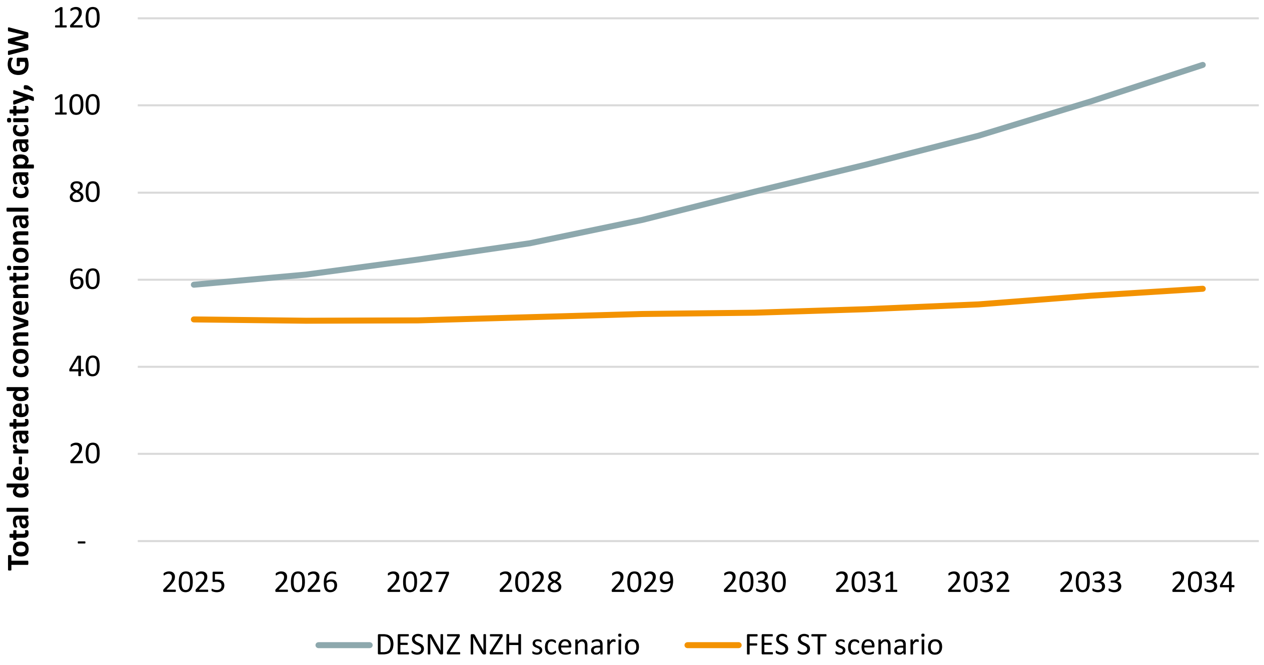
- **Option 1: LOLE combined with Loss of Load Duration (LOLD) –** both the literature review and the first part of the modelling have identified that in certain scenarios, as we decarbonise the LOLD could increase over time. This means stress events lasting longer and potentially introducing additional risks to security of supply. As discussed in the literature review section, a combined metric could be a useful way to manage different types of security of supply risk, so the LOLD and LOLE metrics are used in combination in this option.

Option 0: LOLE

To understand the impact of alternative reliability standards, we first need to understand how much de-rated conventional capacity (defined as not intermittent, therefore including interconnectors) would need to be available under the current reliability standard – a LOLE of 3 hours. Conventional capacity is presented as this has typically been the procurement focus of capacity mechanism schemes. It is worth noting that intermittent capacity will also contribute to capacity adequacy in the future, but it is typically procured through a dedicated support scheme. Some conventional capacity may also be procured outside the dedicated capacity mechanism scheme (e.g. Contract for Difference for Hinkley Point C nuclear plant) and some may have already been procured (e.g. significant volumes of new-build capacity has already received 15-year agreements in the Capacity Market and is therefore set to deliver past 2034).

The chart below shows that the total de-rated conventional capacity necessary for maintaining the reliability standard (LOLE of 3 hours). For the 2022 DESNZ Net Zero Higher Demand scenario, this shows the total de-rated conventional capacity increases considerably over time from 59GW in 2025 to 109GW in 2034. This increase is in line with total and peak demand increases in the 2022 DESNZ Net Zero Higher scenario. In comparison, the total de-rated conventional capacity increases more slowly in the 2022 FES ST scenario with only modest increases from 51GW in 2034 to 58GW in 2034. This reflects the smaller increases in both peak demand and overall demand in this scenario. The results for the key identified metrics for these scenarios with the current reliability standard can be seen in the section above.

Figure 38: *Total de-rated conventional capacity over time and between scenarios (including interconnectors but excluding storage plants)*



An additional area affected by the reliability standard is the derating factors of intermittent renewables (wind and solar) and limited duration storage. These derating factors are calculated using the Equivalent Firm Capacity (EFC) methodology, described in further detail earlier in this report.

For intermittent technology types (wind and solar), the average EFCs show a small decline over time in both scenarios as the overall capacity of these technologies increases. For example, EFCs for offshore wind decrease from 17% in 2025 to 14% in 2034 in the DESNZ NZH scenario. This means that each additional unit of intermittent renewables makes marginally less contribution to security of supply thus pulling down the average EFC as more of these technologies are added to the system. The average EFCs are slightly higher in the DESNZ NZH scenario compared to the FES ST scenario. This is due to intermittent renewables making up a larger proportion of the system in FES ST, and therefore the marginal contribution of additional intermittent renewable capacity being smaller.

The average storage EFC represents the EFC for the whole storage fleet so includes all durations of storage in the measure. This value is lower in the FES ST scenario than in the DESNZ NZH scenario due to proportionately higher amount of storage present in this scenario. While the EFCs go down over time as additional storage capacity is added to the system in both scenarios, the drop is less pronounced in the FES ST scenario due to more of this being additional long-duration storage (for example pumped storage).

Figure 39: *Average Equivalent Firm Capacity (EFC) of plants with different technology type over time in the DESNZ Net Zero Higher scenario*

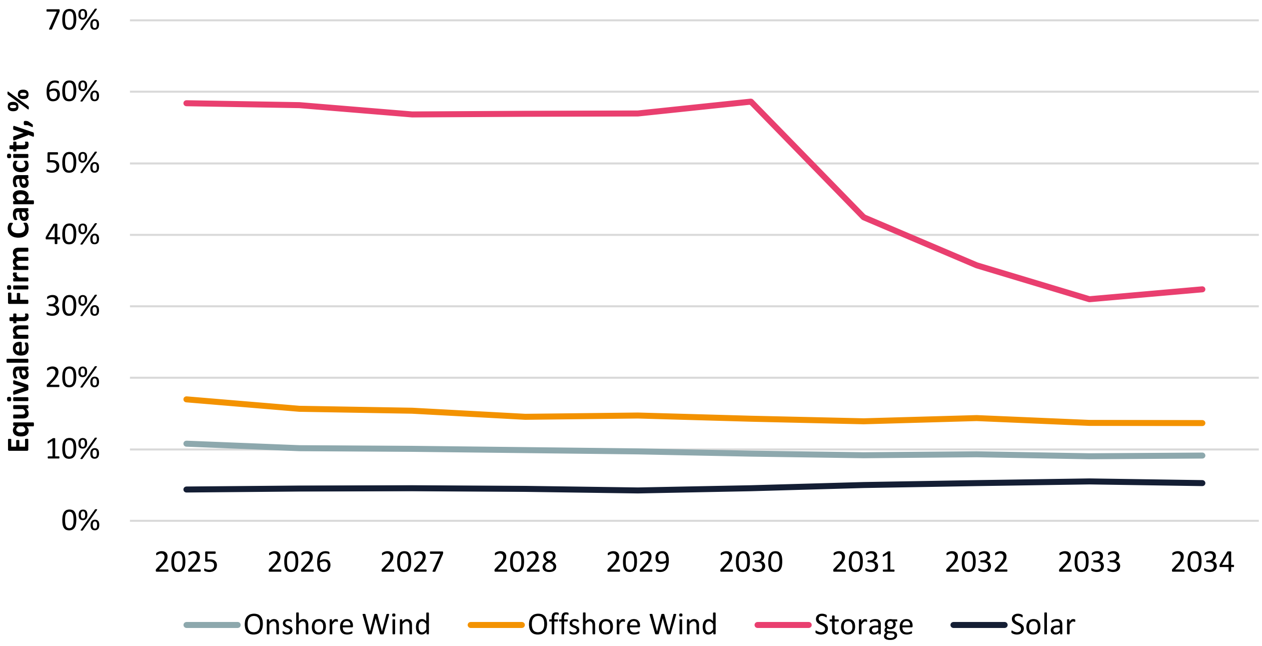
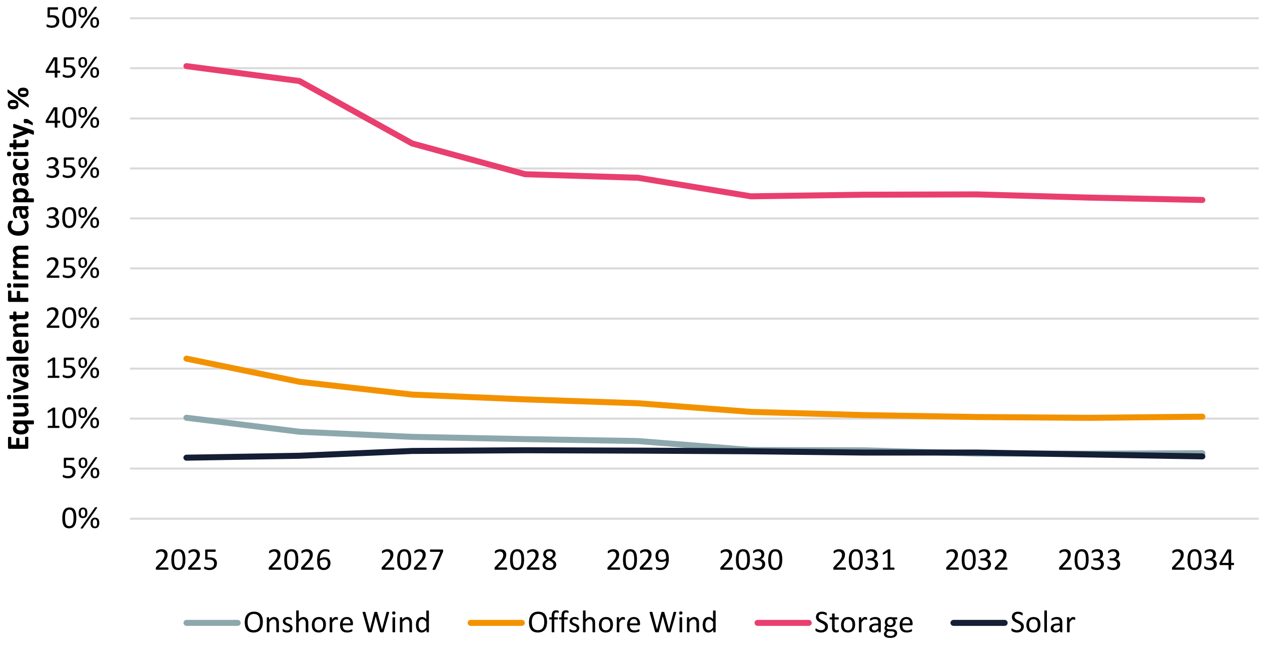


Figure 40: *Average Equivalent Firm Capacity (EFC) of plants with different technology type over time in the FES System Transformation scenario (baseline with the reliability standard at LOLE of 3 hours)*



Option 1: LOLE combined with LOLD

Longer stress events may represent additional costs to consumers compared to what was assumed as part of the reliability standard setting process. The LOLE of 3 hours/year was derived with a set of assumptions about the length of events, and if the actual length of a typical event increases, this value may no longer be suitable. A combined metric can address this, by ensuring that the system satisfies both the existing reliability standard of 3 hours/year (LOLE) and that the expected length of a single event (LOLD) also does not rise above a certain level.

For this analysis a LOLD target of 3 hours/event is used because it is the level of LOLD in 2025 in the DESNZ NZH scenario. In reality, the LOLD target could be informed by further studies incorporating VoLL, taking into consideration the value that consumers place on limiting the length of a single system stress event. As the LOLD only changed more significantly over time in the FES ST scenario (as discussed in 5.1) the LOLE combined with LOLD reliability standard is only implemented for this scenario. This is done by taking the system with a LOLE of 3 hours/year in a given year, and then adding additional, uniform de-rated capacity until the LOLD metric reaches 3 hours/event.

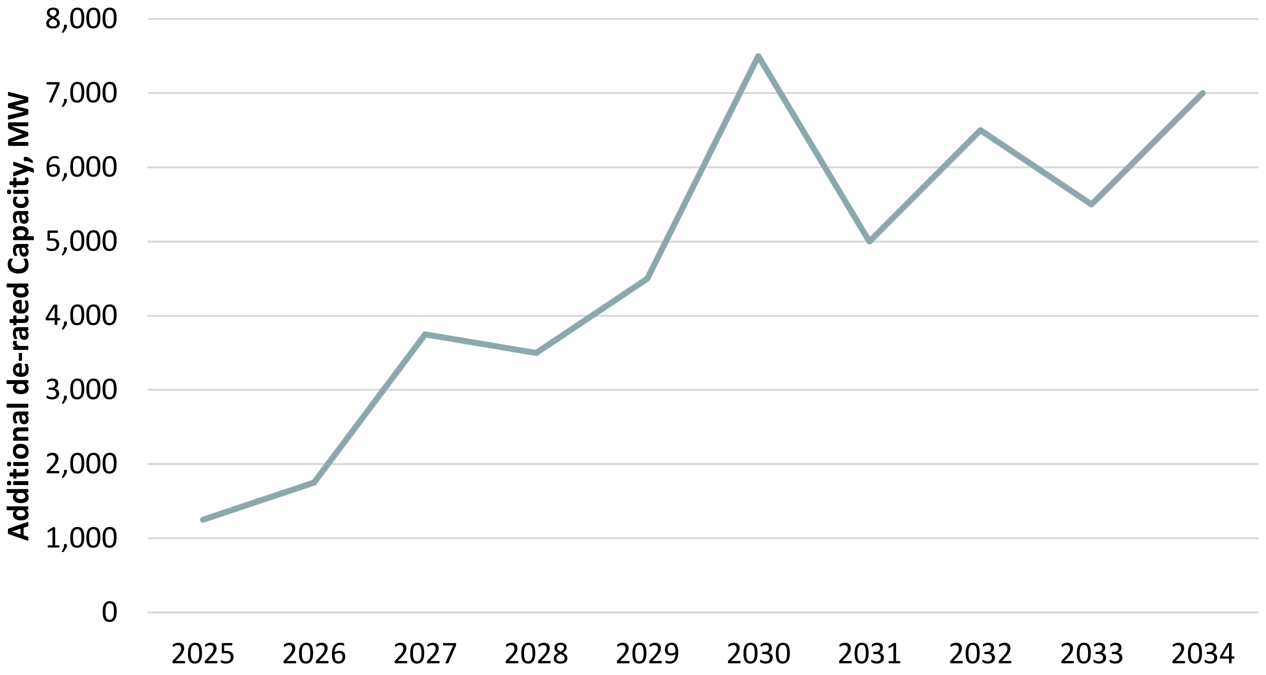
The amount of additional capacity needed to maintain the average length of an event at 3 hours/year in the FES ST scenario increases over time from a low of 1,250MW in 2025 (2% increase) to 7,000MW in 2034 (13% increase) as shown below. This is in addition to the total de-rated conventional capacity that needs to be available to maintain the reliability standard of LOLE of 3 hours/year, as described in Option 0 above. Maintaining this additional capacity would represent additional costs to the system[26]: using the net Cost of New Entry (net-CONE) currently in use in the Capacity Market, £49/kW (real 2013), these costs would range from £61 million in 2025 to £343 million in 2034. (Noting that the current net-CONE is based on the cost of a new OCGT plant, set in 2013).

A more detailed consideration of the total system and consumer costs of this additional capacity falls outside the scope of this project but changing the reliability standard could have a considerable impact. For example, while capex costs are higher due to higher levels of build through the capacity, higher total generation availability would reduce the frequency of tight events in the wholesale market, which could lower consumer costs, and reduce the system costs due to reductions in EEU. Changing EFCs for storage could also enable these to gain more revenues and be a more economically viable investment increasing storage capacity on the system and impacting generation costs with more storage on the system enabling storage discharge to gas generation.

---

[26] This cost may fall into different areas within the normal system costs framework used by DESNZ. Additional OCGT capacity cost would increase capex costs and thus overall system costs. Consumer costs would also increase due to increased CM payments to these plants. However as noted below, benefits may also come through in different ways a as a result of the system being more secure

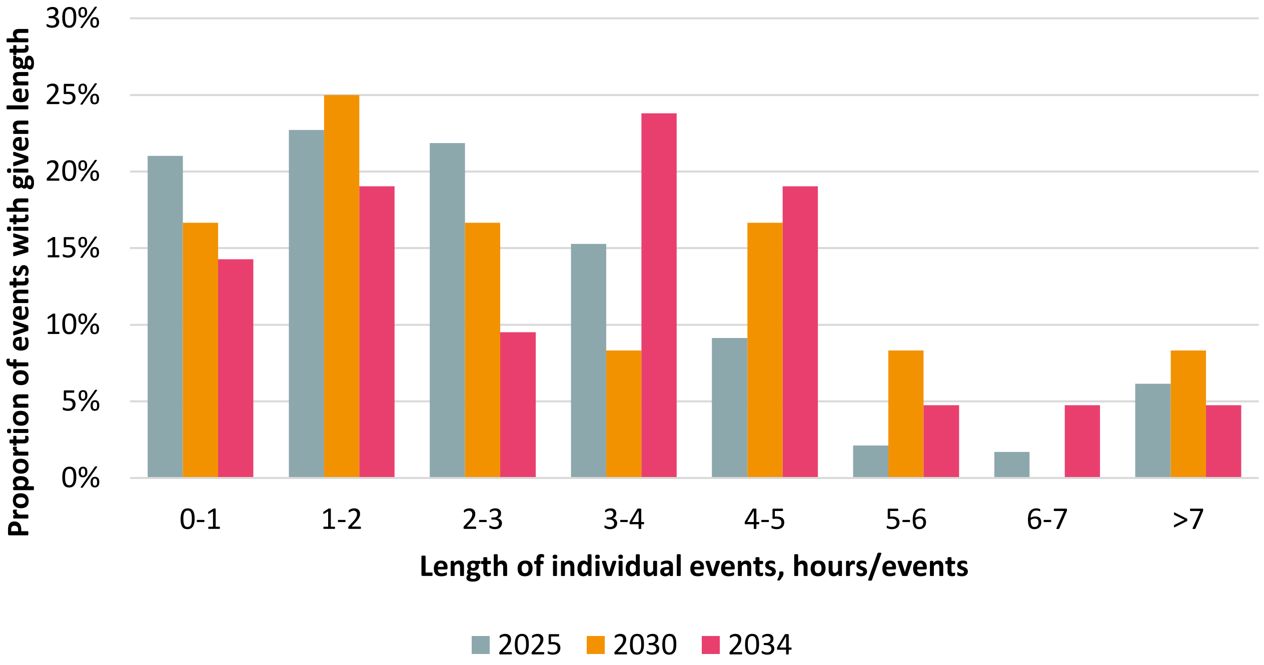
Figure 41: *Baseline de-rated conventional capacity and the additional de-rated capacity needed to maintain the combined reliability standard in the FES ST scenario time compared to the current reliability standard*



As a result of the changing reliability standard, the results for the other security supply of metrics outlined in 4.1 also change. In general, most metrics are lower with the combined metric reliability standard compared to the LOLE-only reliability standard. Overall, this suggests the combined metric of LOLE and LOLE leads to a more secure system. However, it is important for policymakers to decide the value of this extra security that this approach provides against the additional costs of procuring more capacity through the CM as a result of the change.

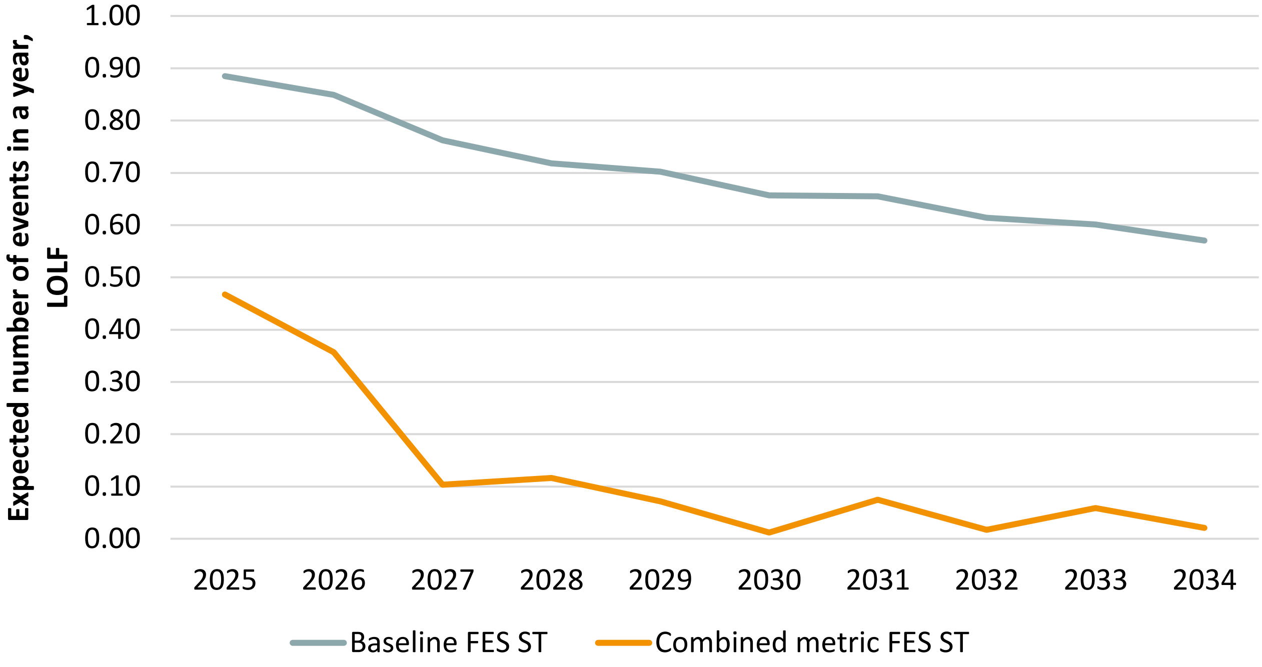
With the average LOLD fixed at 3 hours across the simulations, the distribution of the LOLD changes substantially compared to the original FES ST scenario (shown in figure 14 above). The proportion of simulations where events are longer than 5 hours now remain low throughout even in 2034 with more shorter events at 5 hours or less. This means that the more extreme longer system events that were more likely in the original FES ST scenario are now less likely, so the combined metric is also controlling the higher volatility aspect of the scenario. It is worth noting that events shorter than 3 hours still decrease over time despite all the additional capacity on the system, likely due to the additional storage capacity in later years leading to longer events, as discussed above.

Figure 42: *System event length distribution (LOLD) over time, FES ST Scenario with a combined LOLE and LOLD reliability standard*



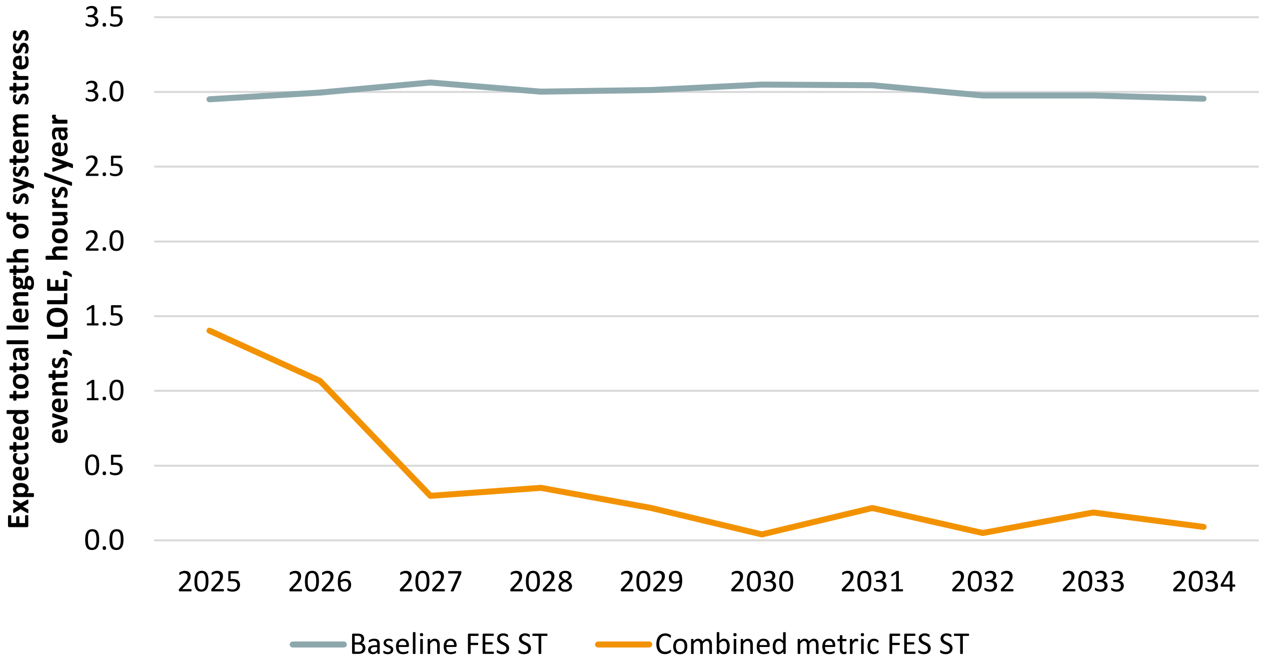
In addition, the Loss of Load Frequency is considerably lower when the combined metric is used and it is progressively decreasing over time, as additional capacity is added onto the system. This is because additional firm capacity is available and this reduces the probability that supply and demand get closer, independently of the impact on the length of an event. This shows that targeting both LOLE and LOLD will decrease the frequency of events in addition to the length of events. It is worth noting that LOLF reaches almost 0 in some years, as the increase in the duration of an event due to the additional storage capacity over the years means that the volume of added capacity needs to be so high to maintain LOLD of 3 hours/year that there are almost no events taking place.

Figure 43: *Expected number of system stress events (LOLF) in a year, FES System Transformation scenario with LOLE only reliability standard (Baseline) and LOLE/LOLD combination (combined metric)*



Correspondingly, the LOLE also decreases over time. This is because as the average duration of an event (LOLD) stays at the target level of around 3 hours while the frequency of an event occurring decreases.

Figure 44: *Expected total length of system stress events (LOLE) in a year, FES System Transformation scenario with LOLE only reliability standard (Baseline) and LOLE/LOLD combination (combined metric)*



Finally, the EEU also drops over time, reversing its upward trend in the baseline scenario. This is because the drop in LOLE is so high that it counteracts the overall increase in peak demand, as seen in the normalised EEU graph. The same drop is observed for the 95th percentile of both the unserved energy and the normalised unserved energy, as well.

The behaviour of the unserved energy, and especially the normalised unserved energy is of particular interest, as these metrics increased in both baseline scenarios. All of the metrics react strongly to using LOLD as a combined metric with LOLE, with the previously observed increases reversing, and the values dropping well below the values observed today, in the same way as has been seen with LOLE and LOLF.

As the combined metric was practically implemented by adding additional generation capacity to the system to keep the LOLD value at the current value, we can also infer what would have happened if the normalised EEU metric had been chosen as the alternative metric to use as a combined metric with LOLE. As the capacity added to maintain LOLD was enough to cause the normalised EEU to decrease, the capacity to maintain the normalised EEU would necessarily have been lower. In turn, this also means that LOLD itself would have continued to increase from the present value.

Overall, this means that LOLD functions as a strong alternative metric to be used in combination with LOLE, as all the other summary statistics associated with stress events stay stable or decrease when using it, while using the normalised EEU would have meant that some aspects of system stress events, namely their average duration, would continue to worsen compared to where they are today. This provides further justification for using the LOLD as a combined metric in this study.

Figure 45: *Summary statistics of unserved energy over time, FES System Transformation scenario with LOLE only reliability standard (Baseline) and LOLE/LOLD combination (combined metric)*

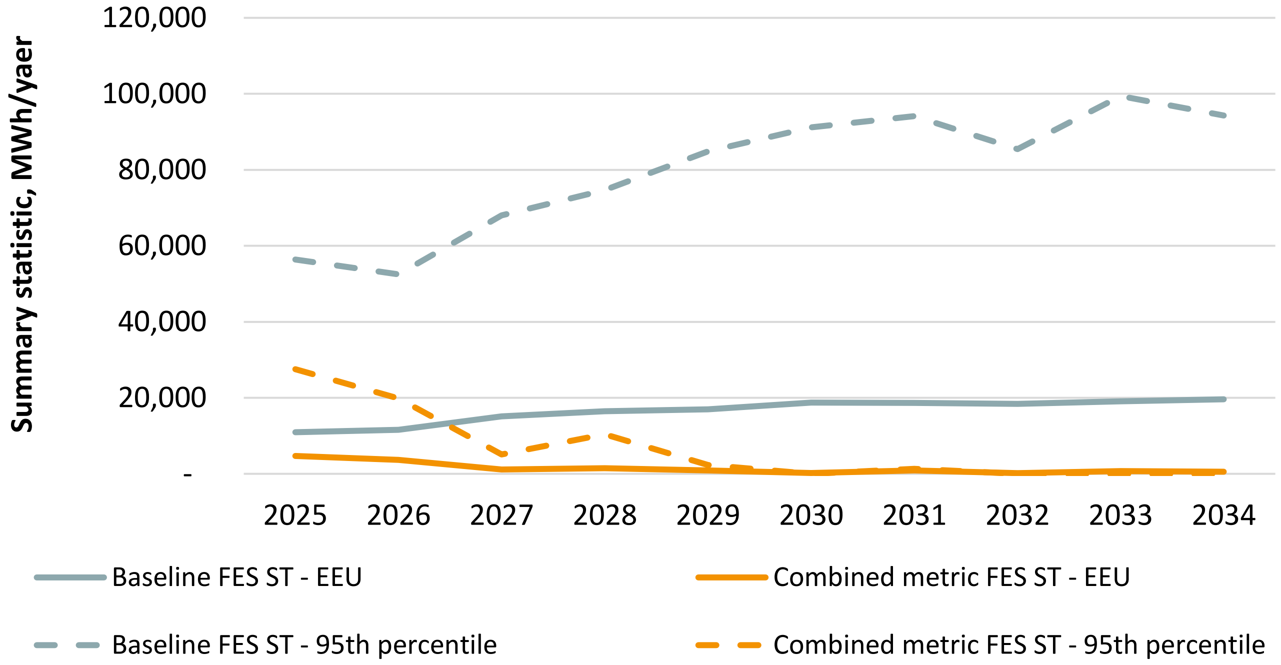
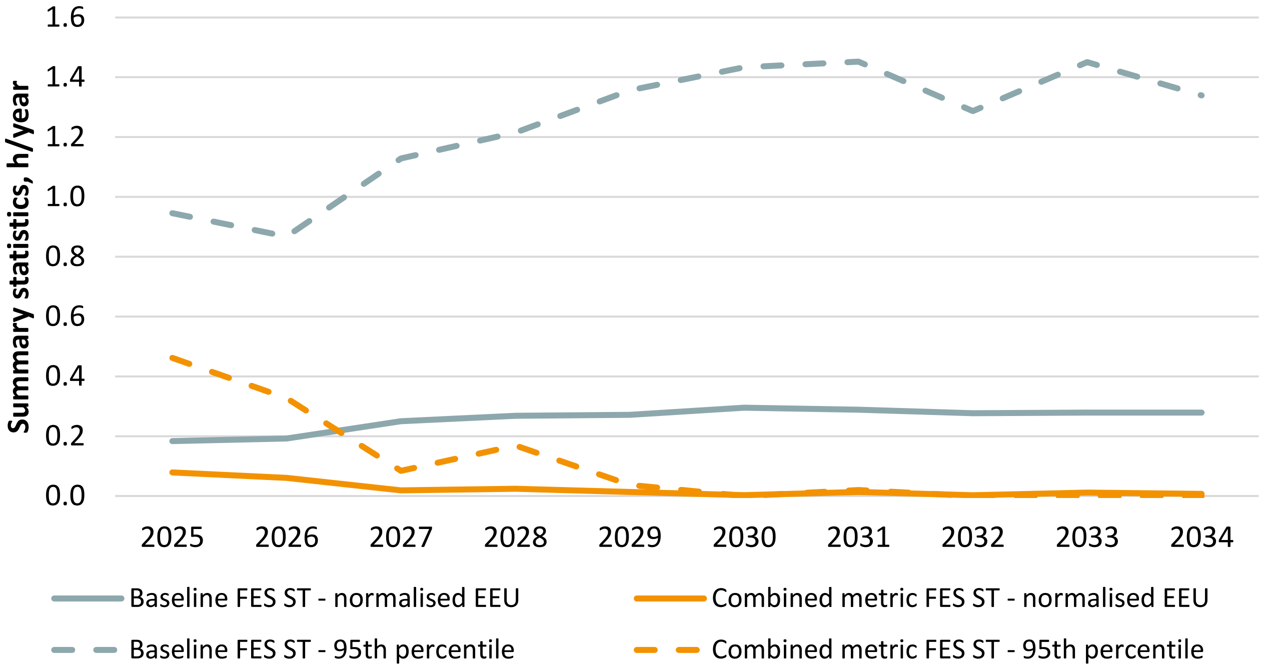


Figure 46: *Summary statistics of normalised unserved energy over time, FES System Transformation scenario with LOLE only reliability standard (Baseline) and LOLE/LOLD combination (combined metric)*



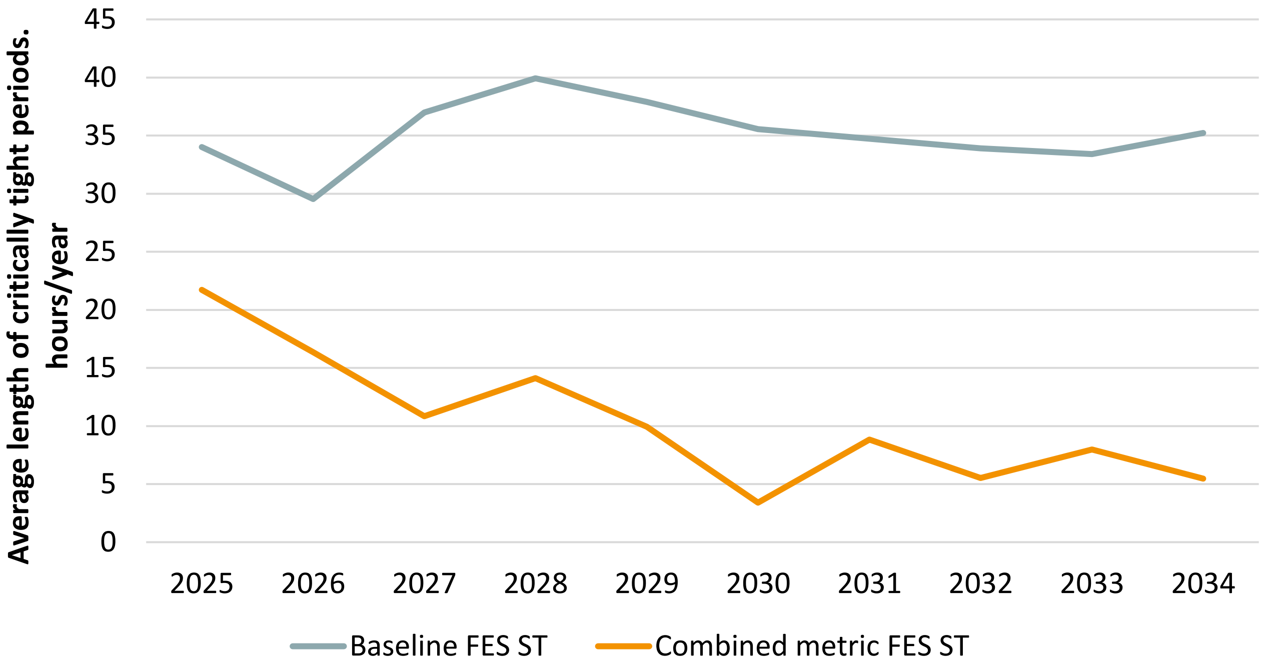
Introducing a revised reliability standard metric in practice requires the consideration of two additional aspects that largely fall outside the scope of this study: economic allocative efficiency and practical implementation in a government subsidy scheme. LOLE has a special place among the metrics due to its use as an economically allocative efficient reliability standard, as described in the literature review section above. Under such theory, if the LOLE is in practice lower than 3 hours/year, then the system is over-secured, and consumers are paying for additional capacity that they may not value. This would be the case if the additional LOLD metric were introduced, as LOLE is well below the 3hours/year, particularly in later years. However, there are two additional considerations to take into account.

First, the existing reliability standard was derived with the assumption that the nature of typical events will not be changing over time. This will need to be revisited as the nature of system events are changing when estimating the level of capacity consumers are willing to pay for, as described in the section on VoLL in the literature review section. Second, LOLE values in the GB power system have been considerably below 3 hours/year in recent years, despite this being set as a reliability standard, due to the target-setting process of the Capacity Market mechanism.

This process, known as the ‘Least Worst Regret’ (LWR), explores a number of different but feasible scenarios and chooses the value whose worst regret across the scenarios is lowest. As the regret cost of a high number of system stress events is typically higher than the regret cost of unused capacity, this process will often result in a degree of prudence towards procuring additional capacity and if the reality is more positive than the scenario that the capacity was procured for, the LOLE value will also be lower. Therefore, the scheme may often lead to lower LOLE values than the reliability standard in practice. This would mean that the additional LOLD target could potentially be achieved without setting it as an explicit target in some cases. However, this analysis does not consider the exact scenarios in which this relationship holds and the LWR approach is designed to capture a specific type of risk than an additional LOLD target would.

Critically tight events also decrease under the combined LOLE/LOLD reliability standard. This is due to the increase in capacity procured leading to the generation curve being shifted upwards in many periods which directly reduces the time periods where supply is close to demand. Keeping in mind that critically tight periods are affected by the storage algorithm included in the model, this implies that there would be additional capacity on the system and there would be fewer periods of scarcity pricing, reducing overall system costs.

Figure 47: *Expected total length of critically tight events in a single year, FES System Transformation scenario with LOLE only reliability standard (Baseline) and LOLE/LOLD combination (combined metric)*

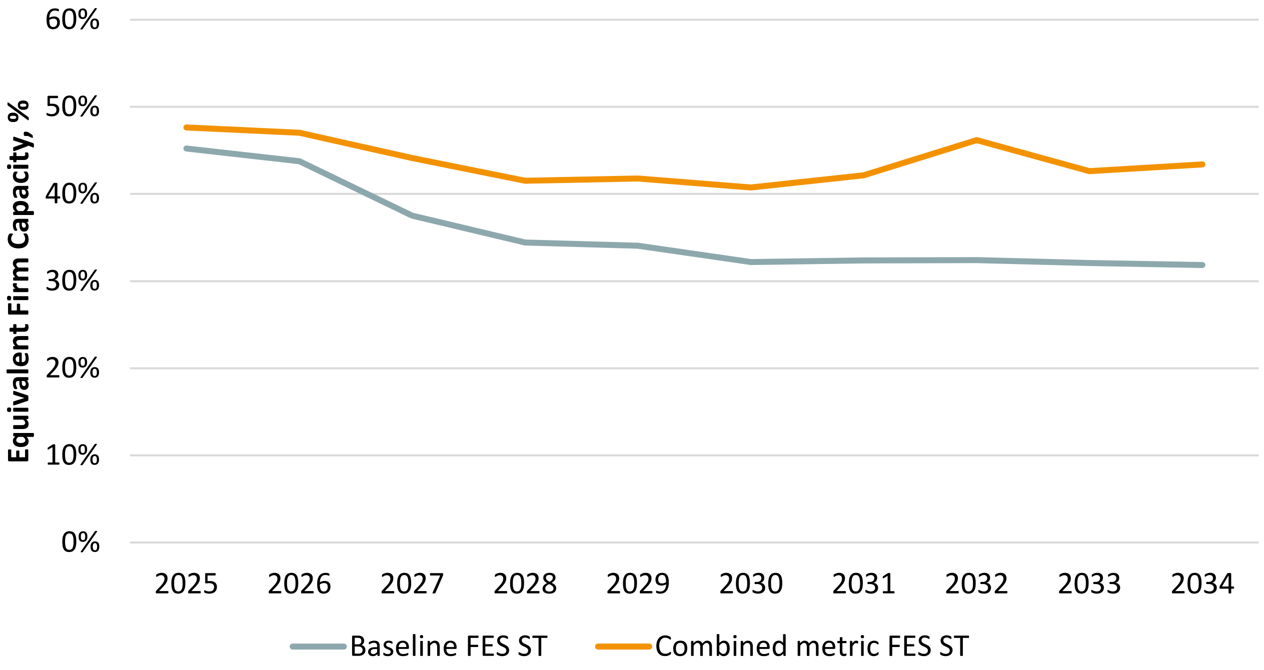


In addition to changing the amount of capacity that needs to be procured, changing the reliability standard may also impact the technology mix present in the Capacity Market (CM). In this study, we have modelled the impact on Equivalent Firm Capacities (EFCs), which are used to calculate derating factors for intermittent generation and for storage. As there is little change in EFCs for intermittent generation technologies against the baseline, and modelling the full impact on the technology types procured through the CM is out-of-scope for this study, we focus on the insights from this analysis relating to the impact on limited-duration storage plants.

Shorter system events mean that storage plants can contribute to negating them more frequently. This means that the EFC of storage plants stays stable over time in the scenario that uses the combined metric and is at higher levels than in the baseline scenario, as it has a higher contribution from storage plants.

Derating factors for most other technologies (other than interconnectors) are derived directly from historic availability during tight system periods, and therefore they will not be directly affected by the change in the reliability standard. However, it is worth noting that as the total volume of some technology types increases, such as Demand Side Response, this derating factor calculation methodology may change in the future.

Figure 48: *Equivalent Firm Capacity (EFC) of storage plants, FES System Transformation scenario with LOLE only reliability standard (Baseline) and LOLE/LOLD combination (combined metric)*



Overall, the analysis shows that a combined reliability standard of LOLE and LOLD would provide additional capacity adequacy leading to no change or a reduction in all key reliability standard metrics in the FES ST scenario. This is a result of this reliability standard requiring the procurement of up to 16% additional derated capacity compared to the baseline conventional capacity already present to meet the existing reliability standard metric. Procuring this additional capacity comes at additional costs but may also provide additional system benefits. The direct costs of procurement in the Capacity Market can be estimated based on net-CONE currently in use in the Capacity Market: this additional capacity would cost an additional £343m in 2034. This reliability standard could also have the additional benefit of higher storage EFCs which could help support the flexible capacity that the system needs as we decarbonise, as well as a decrease in wholesale market generation costs due to higher total available generation capacity.

4.3. Possible further analysis

DESNZ may want to undertake additional work to gain a fuller understanding of the implications of changing the reliability standard to inform wider capacity adequacy reform policy development. Through the literature review and modelling undertaken in the study, Frontier and LCP have identified some further modelling and analysis that could be completed which is out of scope of this study:

- **Alternative scenarios** – The results from the modelling have shown that the make-up of the future power system is an important factor in how security of supply metrics will change in the future. Modelling results have shown key differences between the 2022 FES ST scenario and the 2022 DESNZ NZH scenario, particularly the shape of daily demand. As a result, we would recommend looking at other scenarios with a particular focus on how the shape of demand will change in the future and the role of flexible technologies within that. For example, in this study some assumptions had to be made on the way that daily demand changes in the FES scenario so more detailed assumptions on this from NG ESO could change the results.
- **Review the target level for LOLE –** As the system changes, LOLE may be retained as the future reliability standard, but its current target of 3 hours/year may no longer be fit for purpose. A refresh of analysis previously completed[27] to determine the LOLE target determined as the ratio of the Value of Lost Load (VOLL) and the Cost of New Entry (CONE) may be needed.

  For example, as total demand increases with the power system becoming more demand-side driven and digitalised, the VOLL may change and therefore the economically appropriate LOLE to be used as a reliability standard may change. If the resulting new LOLE is lower than the existing value, implying a lower appetite for system stress events than currently assumed, the impact due to the nature of system events would be expected to be limited. However, if a larger LOLE value is returned, implying a higher risk appetite, the nature of system events may change more significantly, as system stress events may more often be driven by low generation periods (such as a wind drought) leading to the entire daytime demand being unmet (excluding the night). This could lead to more extreme, longer stress events on average.
- **Target level for a LOLD metric** – In this analysis, we assumed that the LOLD target for the combined metric approach was set at the 2025 level (from the DESNZ NZH scenario). Determining its level based on VOLL, or an assessment of alternative analytical frameworks for setting the level of a reliability standard beyond VoLL was out of scope of this project. Exploring what type and level of VoLL would make it worth including LOLD as a combined metric could be a useful exercise to undertake.

  One approach may include two studies where one follows from the results of the other. An initial study could be used to understand the current relative risk appetite of consumers in Great Britain – for example, consumers may have a high tolerance for more overall events

---

[27] Annex_C_-_reliability_standard_methodology.pdf (publishing.service.gov.uk)

as long as each event is shorter. This initial study would therefore identify some potential, more complex VoLL relationships that take the length and frequency of events into account. These relationships could be used to gauge whether there is a need in the population to consider a second reliability standard – for example, if the results indicate that consumers only care about the total length of events in a year, rather than the frequency or the length of events, then the current approach of using LOLE by itself would be appropriate. The second study would then be able to focus on estimating the value of VoLL as described in the relationships identified in the initial study.

- **Use of Critically Tight Periods** – Critically tight periods are useful in characterising the true costs and benefits of a system using to explore the impact of using different reliability standard metrics. The model could be expanded by letting storage plants dispatch during critically tight periods to provide a more accurate metric on critically tight periods. Any remaining critically tight periods could then be included as an additional cost for a given reliability standard metric.

  Alternative approach to calculating capacity shifts – To arrive at the exact level of LOLE and LOLD that was specified in a given scenario, we shifted the level of firm capacity available. NGESO uses the same approach when they run the UEM to derive target capacities for the Capacity Market auctions and we are using shifts that are comparably sized (albeit somewhat larger). The sizes of the necessary shifts are similar between the two main tested scenarios (DESNZ NZH and FES ST). The alternative approach is to add and remove individual power plants until the LOLE and LOLD levels of the system reach the pre-specified values. This returns a more realistic scenario with the outage behaviour of the removed and added power plants feeding into the expected system stress event – it is possible that this approach would lead to higher additional capacity needed to maintain the LOLD level of 2025 over the years, at higher overall cost. In practice, the type of power plant that are added and removed would have an impact on the outcome – this would be different in different scenarios and would likely change as the power system evolves and the market environment changes, however when adding/removing peaking plants with high derating factors (>95%) the difference is unlikely to be significant.

This publication is available from: www.gov.uk/government/publications/review-of-electricity-market-arrangements-rema-technical-research-supporting-consultation