Considerations in support of the 2022 Reliability Standards and Settings Review

Briefing Note prepared for the Australian Energy Market Commission

Pierluigi Mancarella

The University of Melbourne

August 2022

Table of Contents

Executive Summary

This document is a briefing note prepared for the Australian Energy Market Commission (AEMC) with regards to the 2022 Reliability Standards and Settings Review (RSSR) that is being conducted in the context of the Australian National Electricity Market (NEM).

The main conclusions and recommendations in the context of an evolving power system dominated by variable renewable energy (VER) supply and energy-limited resources (ELR) are the following:

- The **reliability risk profile of future systems will be fundamentally different** from what witnessed historically, as the main underlying parameters are no longer approximately normally distributed, temporally uncoupled, and independent;
- Consideration should be given to fundamental gaps and drivers for changes in existing reliability thinking, particularly with regards to weather-driven aspects such as **ramping**, **dark doldrums**, and **increasing physical and market variability and volatility** (**and therefore risk**);
- **Capacity margins**, loss of load probability (**LOLP**), and loss of load expectation (**LOLE**) metrics may be (very) **unsuitable** in future systems;
- With **energy adequacy**, as opposed to capacity adequacy, becoming more relevant, integral metrics such as **energy not supplied** (ENS) – or equivalently unserved energy (USE) – are **much more suitable** than others;
- **Inclusion of multiple metrics** should be considered to fully capture the changing risk profile: it isn't eventually a question of which one is really best, as they all have different and important messages;
- **Introduction of tail metrics** (e.g., based on the conditional value at risk – CVaR) or other approaches might be essential to capture risk moving forward, especially with more volatile climate-change driven events across different years;
- This is especially true in a context of **risk aversion for both consumers and investors**;
- **Risk-aware decision making** could replace classical cost-benefit analysis, for example through **weighted averaging of expected and tail metrics** or least-worst weighted regret (**LWWR**) analysis;
- In fact, the changing system characteristics based on VRE and ELR will manifest more and more into **operational risk and market volatility** (particularly year by year, particularly in the context of climate change), which also brings **significant investment risk** if not carefully hedged through suitable incentives;
- As **demand** becomes **more and more elastic** considerations for capturing its role in the reliability/economics interplay should be given, for example, by **introducing price thresholds that capture near-outages event periods** which could augment outage-based reliability metrics in order to fully capture system stress conditions;
- In the next future, **extreme price volatility or technical measures** (e.g., based on area control errors or operational reserve scarcity) could also be used to **signal resource inadequacy that may be associated with issues such as lack of flexibility**, which could be dealt with in planning and within the reliability standards remit, as opposed to a more traditional lack of capacity/energy;
- A **new form of reliability standard** that could better address the emerging risk in low-carbon grids could be of ***composite*** nature, for example, **a linear combination of expected value and tail value measures of a given metric**, e.g., for the energy not supplied metric, a weighted

combinations of expected energy not supplied and a predetermined CVaR level of the energy not supplied.

- The relevant **settings and parameters informing a composite standard** could be determined by different types of **parametric studies** looking at the system's cost-reliability performance under different scenarios for more "average" and "rarer" events, potentially considering **budget constraints** that would **indicate the willingness to pay for an "insurance premium" against extreme events**.
- Due to the disconnect between the risk-aversion attitude of different agents and the complexity in mapping risk-aware reliability objectives into market incentives, it should be recognised that **markets alone might not be able to deliver on the requirements for extreme events, thus justifying the rationale of different types of interventions** such as strategic reserves, government-backed investments, and so forth.
- Such **insurance policy mechanisms** to deal with events of unique nature such as high impact low probability ones **should be efficiently coordinated with market operation and price signals** and **enhance the scope and performance of a market,** *and not replace it*.
- Moving forward, in order to inform future standards **significant modelling efforts** are needed to fully capture the risk of a much more weather-dependent power system and the operational complexity of variable and energy-limited resources, including probabilistic studies with detailed representation of time-ahead and real-time operation of the system.

Contact: pierluigi.mancarella@unimelb.edu.au

1. Introduction and outline

This document is a briefing note prepared for the Australian Energy Market Commission (AEMC) with regards to the 2022 Reliability Standards and Settings Review (RSSR).

However, in order to allow unconstrained thinking, we will take a perspective that is as general as possible and looks at the fundamentals of reliability, the way that standards and settings may be defined, and their role, rather than at any specific reference to the existing regulatory environment.

To inform this note, we looked into the international state-of-the-art (academic publications, reports by government agencies in different countries, international projects, etc.) and various discussions and outputs from international working groups in professional societies, including IEEE and Cigre, or research organizations such as the Energy System Integration group (ESIG) or the Electrical Power Research Institute (EPRI). In particular, much of the recent thinking on what is Australia is referred to as “reliability”, which internationally is more consistently referred to as “resource adequacy”, has been informed by ESIG activities, in which the author is actively involved.

A nice summary of the challenges and policy options associated with reliability moving towards more weather-based systems is provided in Figure 1[1], sourced from ESIG and the Global Power System Transformation Consortium (GPST). In particular, it may be appreciated how in the longer term the main physical reliability risks are associated with the uncertainty that climate change brings to future system impact and requirements, as well as the simultaneous retirement of traditional generation units that have been the historical providers of security and reliability. On the other hand, several mitigation options may become available for policy makers, particularly with regards to increasing load flexibility in the medium term and facilitating investment in both new generation and network technologies in the longer term. Of course, while suitable design of market and regulatory framework and settings to enable these investments is a key option for physical risk mitigation, how to do this optimally in a fast-changing technology and political environment also brings several challenges, some of which are outlined in these notes.

[1] Energy Systems Integration Group and Global Power System Transformation Consortium, “Ensuring Not Only Clean Energy, but Reliability: The Intersection of Resource Adequacy and Public Policy.” November 2021.

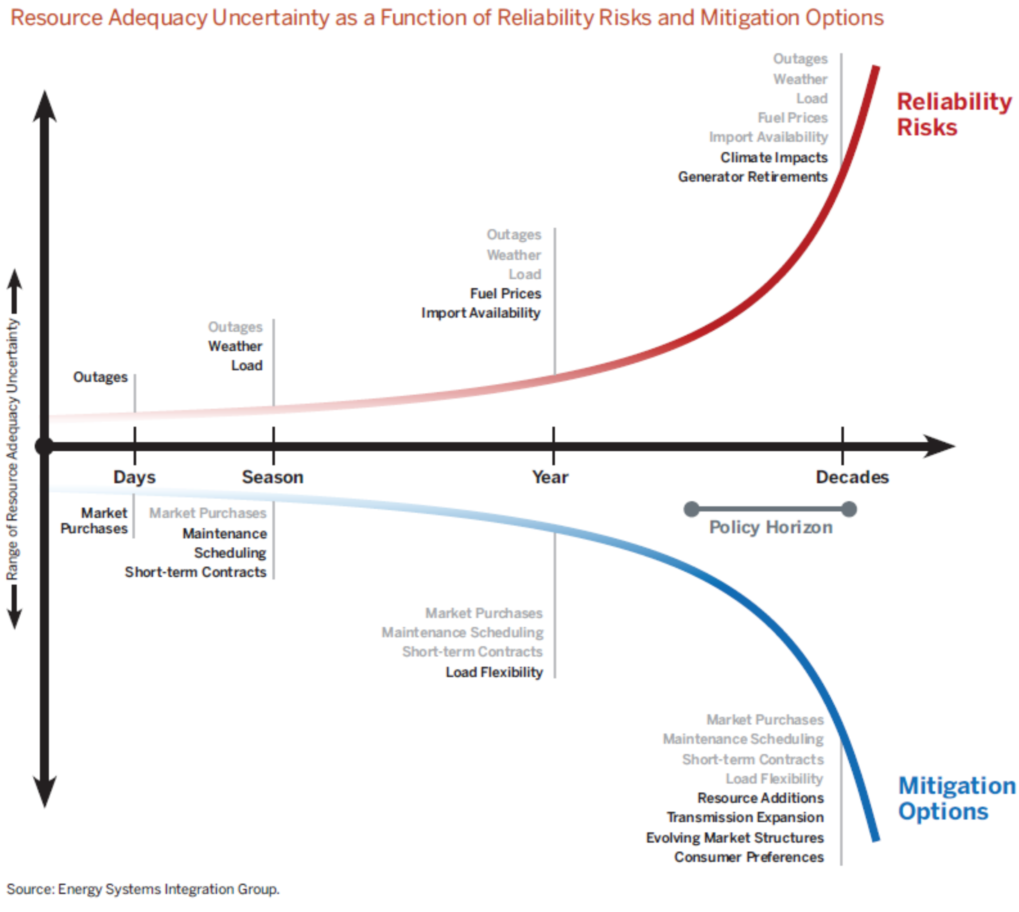


Figure 1. Reliability challenges and mitigation options for policy makers and regulators (Source: ESIG and GPST[1]).

The document is organised as follows. Section 2 provides general definitions about reliability, adequacy and security to clarify the terminology used in the context of a changing system. Section 3 explores the fundamental reasons behind and the main features of a shifting risk profiles in resource adequacy, moving from a traditional power system dominated by conventional generators with relatively inflexible demand to a low-carbon power system dominated by weather-dependent variable renewable energy (VRE) supply and different forms of storage and flexible demand (energy-limited resources – ELR), also enabled by the uptake of distributed energy resources (DER). Section 4 provides a critique of and gap analysis in current reliability thinking and metrics and approaches used for its evaluation, including the rationale for incorporating risk-aversion. Section 5 delves into and discusses different options to measure reliability in low-carbon grids, including market impact considerations. Section 6 provides a summary of general considerations and key recommendations for a new, evolved reliability framework that can more effectively deal with the challenges of a low-carbon grid. Section 7 then explores more in detail potential approaches and market implementation issues to deal with new forms of composite reliability standards that can account for tail risk. Section 8 finally contains the concluding remarks and outline potential next steps to inform the Australian regulatory environment.

2. Reliability, resource adequacy and security

Reliability

Internationally, one of the widely accepted definition for *reliability* is as "*a measure of the ability of a power system to deliver electricity to all points of consumption and receive electricity from all points of supply **within accepted standards** and in the amount desired"*[2].

Historically, *to deliver reliability* engineers and regulators have also *practically* distinguished between planning and operation aspects of the physical system and markets, further breaking down the concept into the two major "sub-properties" of *adequacy* and *security*[3].

(Resource) adequacy

Following again Cigre, *adequacy* is *"a measure of the ability of a power system to meet the electric power and energy requirements of its customers within acceptable technical limits, taking into account scheduled and unscheduled outages of system components"*[2]. In this definition, "scheduled" outages usually refer to maintenance-driven outages of generators, transmission circuits, substations, etc., while "unscheduled" outages typically refer to the so-called "credible" contingencies, such as equipment faults, loss of a circuit or a generator, etc.

It is to be noted that in the NEM, a "reliable power system" is defined as one that "*has an adequate amount of capacity (generation, demand response and interconnector capacity) to meet customer needs. This requires adequate investment in capacity, including sufficient investment to cover generator retirements, as well as an appropriate operational framework so that supply and demand can be maintained in balance at any particular point in time*."[4] This definition of "reliability" in the NEM thus effectively aligns with the international concept of "adequacy".

For a system to be "adequate", it needs to be *planned* so that all the required components (generation, transmission, distribution, etc.) are able to supply the expected demand, including considering its forecasted evolution across time (usually a decade, to include the lead time to plan, approve and build transmission infrastructure and major thermal power plants) and including generation and network reserves to allow for scheduled and unscheduled outages.

Security

Security is defined by the CIGRE Working Group C1.27 as "*the ability of the power system to withstand disturbances*", whereby "disturbance" has traditionally and for practical purposes been associated with "contingency"[4]. In turn, a *contingency*[5] can be defined as the "*trip of one single or several*

[2] CIGRE Working Group C1.27, "The future of reliability – definition of reliability in light of new developments in various devices and services which offer customers and system operators new levels of flexibility", TB 715, Jan 2018.

[3] See for example the classical textbook by R. Billinton and R.N. Allan, *Reliability evaluation of power systems*, Springer, 1996. Stoft, in his *Power system economics* textbook[17], also in practice refers to resource adequacy when the discusses reliability standards, though he also elaborates on operational reserve requirements, that are more relevant to security considerations.

[4] According to the National Electricity Rules, Chapter 4, "*power system security is generally concerned with the strength of the electrical power system to tolerate and respond to system disturbances without suffering equipment damage while maintaining quality of electricity supply to customers*".

[5] A *contingency* is, for example, defined by the European Network of Transmission System Operators for Electricity (ENTSO-E) in their "Operational Handbook - Policy 3: Operational Security" as the "*trip of one single or several network elements that cannot be predicted in advance*".

network elements that cannot be predicted in advance". Hence, as mentioned above security is an operational requirement to ensure the lights keep on following contingencies.

In practice, it is adequacy that is most relevant to regulatory considerations associated with reliability, and that is what we will focus on. However, there are *operational* aspects, so in principle more associated with security, that are becoming more and more relevant to adequacy too. In particular, with decarbonized power system, an important aspect to consider are operational issues that might arise from variability and uncertainty of renewables, for example, their ramps and the associated impact on net demand ramps.

3. Reliability issues and challenges in a changing power system

Resource adequacy in conventional systems

The following *assumptions* have historically been made, more or less explicitly, with regards to resource adequacy/reliability studies for conventional systems based on dispatchable generators:

- Focus on discrete outages and credible contingency events for generators and interconnectors;
- Reliability standards thus driven by discrete outage events, modelled with forced outage rate (FOR) or similar approaches;
- Outages assumed independent of each other;
- Weather considered in scope mostly to address annual impact on value of peak demand, with little or no consideration of its impact on the supply side;
- One typical provider of "marginal" capacity, with features that are non-time-variable and independent of the rest of the system.

As a result of the above assumptions, the following considerations may be made to characterise conventional power systems dominated by fuel-based thermal generators[6]:

- Available capacity is relatively constant across time scales (from seconds to years), as it is only affected, mainly with seasonal cycles, by maintenance and forced outages and only to minor extent by weather (in that it my impact the capacity and efficiency of thermal power plants);
- Annual peak power demand is effectively the main driver for reliability requirements, which may be translated into availability requirements for generation/network *capacity* at peak times;
- Energy-limited resources would generally be modelled in terms of their expected available power at peak times;
- Availability of power supply and power demand requirements at peak times could be, in a simplified and approximated way, statistically represented as "normally" distributed based on independent Gaussian probabilities[7];
- For given typical values of generator's FOR, there would then be a direct, easily representable relationship between *installed* capacity and *available* capacity at peak times;
- *Capacity margin* in planning, describing the difference between installed capacity and expected peak demand, could thus directly be linked to operational margins, and suitable rule of thumbs for deterministic reliability standards could also be adopted (e.g., 20% capacity margin);
- All supply shortage events could be described as reliability metric distributions all broadly stemming from the approximated Gaussian representations of supply and demand at peak times;

[6] Hydro and pumped hydro plants have also been historical parts of conventional systems. However, for the sake of simplicity and for the schematic differentiations we are attempting to draw here, we will not explicitly consider them. However, their operational features are indeed basically the same as those of VRE and ELR discussed as part of low-carbon grids.

[7] Strictly speaking, the distribution describing available capacity with conventional generators and based on their FOR parameters could be better described, under appropriate assumptions, as *binomial*. In turn, under conditions that broadly apply to traditional power systems, the binomial distribution of the capacity available at peak times could be approximated by a *Gaussian*.

- The shape of such distributions would be well known and looking relatively similar even in different contexts (essentially, different portfolios of power plants);
- Expected values of capacity shortages would provide key information with regards to their probability distributions too, with standard deviations being essentially linked to FOR parameters;
- An "average" event would thus be a good representation of "most" events under typical conditions (e.g., FOR not too high, say not higher than 10%-15%, and reasonably small uncertainty in peak demand forecast, in the order of up to a few % points).
- Focusing on contexts where ELR penetration would have been limited, power availability at peak times would not be influenced by the past operational history and by any specific control strategy, and forecast uncertainty of system and market conditions would only play a minor role in resource planning and only for a few plants;
- As a consequence, high market prices could effectively signal the need for capacity requirements in real time, if needed, as the power that could be deployed in the market would only be influenced by the available capacity and not by opportunity cost that would have affected past operational strategies and could influence future ones too.

Resource adequacy in low-carbon power systems

New systems are likely to be characterised by the following features:

- It cannot longer be assumed that the available generation capacity in the system is, to large extent, constant (except for outage allowances), as it might heavily change with the weather on a sub-hourly, hourly, daily, seasonal, and yearly basis;
- Different probability distributions are needed to describe weather-driven supply-side power availability;
- Such distributions are generally not Gaussian;
- For example, wind output is often generally described as a Gamma or Weibull distributions, both significantly skewed;
- This means that installed capacities of VRE do not (at all) clearly represent their available power outputs;
- There is also a finite and, in many cases relatively large, probability that the power output might be zero at peak times;
- The available operational margin cannot longer be represented in a Gaussian way, and the resulting distribution of operational capacity margin might be heavily skewed and tailed;
- There is therefore no longer a clear link between *installed* capacity and *available* power output at peak times, and *capacity margins* become completely inadequate to describe system risk;
- The uptake of flexible demand side resources[8], equipped with storage and in case price-responsive, might change the standard deviation of the distribution of the peak demand and even its shape;
- Supply sources are no longer independent as weather may create highly correlated outputs, so that the "diversity" effect that characterises uncorrelated failures of conventional

[8] E.g., batteries, electric vehicles, electric heat pumps with thermal storage, hot water tanks, etc.

generators, which positively contributed to provide for peak capacity support, no longer applies.

- Available supply output from weather-based resources and weather-dependent demand levels become more *anti-correlated*[9], with high likelihood of potentially higher demand due to extremely hot/cold weather and potentially lower supply due to weather-driven loss of efficiency, capacity, outages, etc.
- Because weather-driven supply and demand may have very different seasonal patterns, "net demand" margins should be evaluated beyond the absolute demand peak times and on a seasonal level, for which scheduled maintenance periods of conventional generators might become an important factor too;
- As weather fronts might persist longer than typical peak demands, operational capacity margins might need to be assessed over longer time durations and not only for the maximum peak's times[10]: time-coupling becomes relevant;
- Supply-side ability to contribute to system adequacy is no longer independent of time, with more and critical dependence on the weather characteristics;
- Time-dependence and time-coupling is accentuated by the presence of different types of centralised and demand-side storage and, in general, ELR, which might:
    - Increase the duration of the peak times, thus also indirectly coupling "failure probabilities" of conventional generators and exposing more the system's high demand levels to weather-based supply scarcity (see for example Figure 7 below);
    - Heavily depend on control strategy, system and market forecast ability, past history, and in general anticipated *opportunity cost*:
        - High prices might not be able to steer a response as in conventional systems as ELR behaviour would now depend on volatility-based opportunity cost and inter-temporal operation: this is something to be considered carefully in energy-only markets;
        - Ability of short-term and long-term forecasting of weather, system, and market conditions becomes key, and operational uncertainty influences reliability much more than in conventional systems[11];
- As storage also couples supply and demand, high demand at certain times might lead to less supply being available later, again coupling time instances in the context of reliability;
- Therefore, failure events and distributions of reliability metrics are no longer easy to identify, as they do not stem any more from generally normal supply and demand representations;
- **Key to understand the new risk profile**, the probability distributions of operational capacity margins and associated reliability metrics are likely to be characterised by:
    - Overall flatter shapes;
    - **Much higher year by year volatility;**
    - **Higher skewness and fatter tails**.

[9] This is particularly true with electrification of heating and cooling, which makes demand much more weather-sensitive.

[10] In conventional power systems, time coupling of multiple generating sources that can affect the overall available supply does not normally happen as generation outages are generally independent and mean times to failures are typically at least one order of magnitude higher than mean times to repair.

[11] In conventional systems, operational uncertainty influences reliability insomuch that the scheduling of operational reserves may be insufficient. However, especially with deep penetration of ELR, it becomes essential to be able to properly schedule different resources and reserves in anticipation of (weather-driven) scarcity events.

As a consequence of the changed circumstances in low-carbon grids, it is possible to draw a key set of observations/findings relevant to why existing reliability arrangements may no longer be adequate:

- Supply interruption events will be characterised by greater spread in duration, magnitude, frequency, etc.;
- "Average" reliability parameters may no longer be a good representation of relevant probability distributions, and "tails" that could be important might be overlooked;
- Expected values of reliability metrics may no longer provide a good representation of reliability risk, especially for risk-averse consumers, planners and investors;
- There is no longer a typical provider of marginal capacity, as multiple technologies could now compete for this role;
- As the characteristics of supply interruption events are now more spread, it is not straightforward to identify a "one-size-fits-all" provider of marginal capacity, and a "provider of marginal energy" might also become relevant;
- Supply-side ability to contribute to system adequacy is no longer independent of the system's conditions as the features of weather-based VRE and ELR depend on the rest of the system and the specific control strategies[12];
- Investment in VRE or ELR within different incumbent portfolios might lead to significantly different reliability impacts.

Figure 2 provides a qualitatively illustration of the evolution of the characteristic of the system while it becomes more decarbonised, based on the probability density function (PDF) of total supply and demand, the PDF of the available capacity margin, and the cumulative distribution function (CDF) of the available capacity margin, with indication of the associated LOLP value[13]. It may be appreciated how the supply PDF shifts from Gaussian shape (conventional generation system) to one that resembles more a Gamma distribution (system dominated by renewables), which could be sometimes used to represent wind speed profiles[14].

The effects of the interplay of all the different components discussed above are also illustrated through more illustrative results reported in the Appendix. The results show how the PDFs of different reliability metrics change with the penetration levels of VRE and storage of different characteristics, particularly with respect to an increasing spread of the reliability metrics and to a change in shape. Altogether, the results confirm the importance of capturing the tails of the distributions as well as of possibly using a mix of metrics to have a better understanding of the system risk.

[12] The "*capacity value*" of VRE and and ELR depends on the incumbent system portfolio.

[13] It should be noted that the values and shapes are illustrative only and should be taken as *qualitative* rather than *quantitative*, as they have been designed to highlight key points of the above discussion and not provide numerical results.

[14] For the sake of simplicity, we have ignored the nonlinear effect of typical wind turbine power curves that map wind speed into power output in wind farms, as well as other nonlinearities.

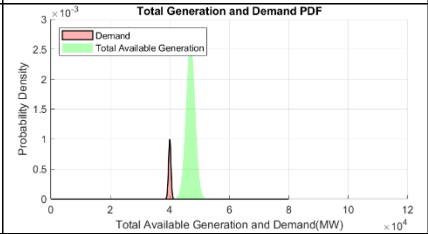
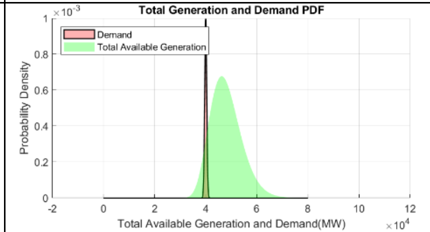
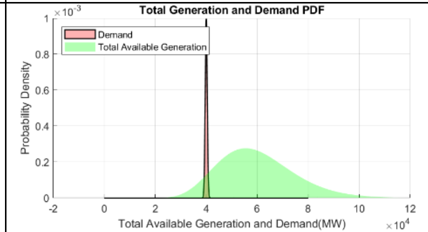
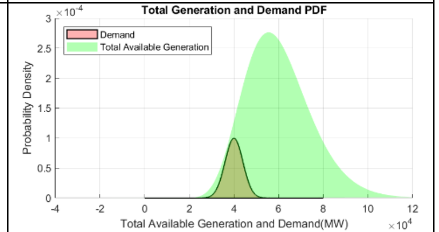
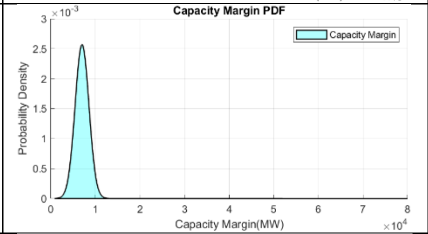
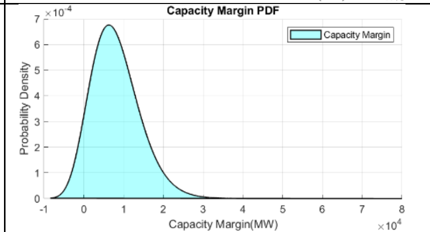
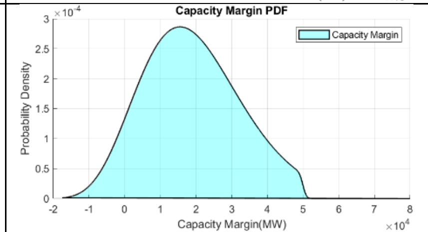
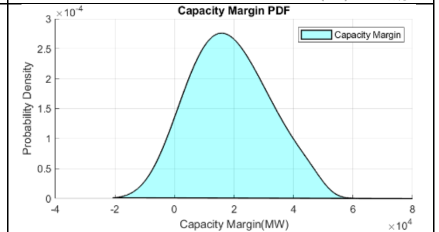
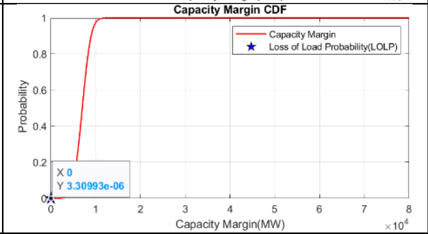
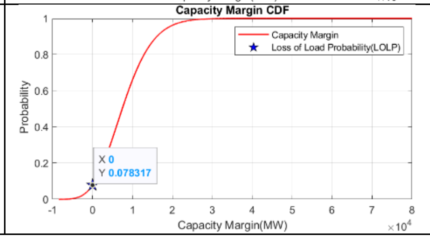
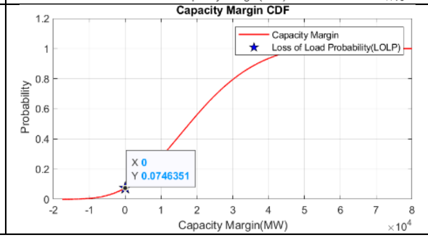
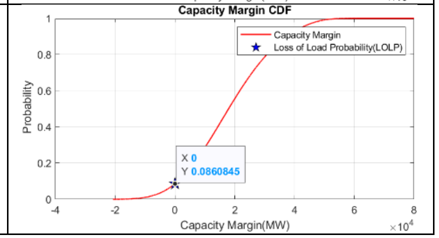
	50GW conventional generation, no VRE, 40GW demand with 1%StD	30GW conventional generation, 40GW VRE, 40GW demand with 1%StD	10GW conventional generation, 100GW VRE, 40GW demand with 1% StD	10GW conventional generation, 100GW VRE, 40GW demand with 10% StD
PDF of total supply and demand				
PDF of capacity margin				
CDF of capacity margin, with LOLP				

Figure 2. Qualitative illustration of changes in shape of probability density function (PDF) and cumulative distribution function (CDF) of supply, demand, and capacity margin, overall describing a changing risk profile, while the system becomes more decarbonized.

4. A critique of and gaps in current approaches to assess reliability and resource adequacy

Typical reliability metrics and standards

Standards may of course be defined in different ways and according to different measures and metrics, for instance, average frequency of outages, expected unserved energy, and so forth. Historically these standards have often been driven by engineering rule of thumbs and practices. For example, in North America and in the UK Loss of Load Probability (LOLP) and Loss of Load Expectation (LOLE)[15] have been, and still are, widely used metrics, with typical standard values in the order of one event in 10 years (LOLP) or 0.1 days per year (LOLE).

Standards may also be defined in the context of a more comprehensive cost-benefit analysis (CBA), as for example done in the NEM where the current UnServed Energy (USE) value of 0.002% was deemed to be the optimal tradeoff between the marginal cost of interruption for customers, valued through the current estimate of the Value of Customer Reliability (VCR) and the investment cost of a marginal generator that could avoid incremental interruptions[16].

Challenges with some "historical" reliability metrics

It should be noted that, although historically developed by engineers, LOLP- and LOLE-based types of standard (may) also have a CBA basis, as discussed by Stoft[17]. More specifically, for the case of inelastic demand, the socially optimal expected duration of load shedding D_{LS}, in hours/year (basically, the LOLE), is given by:

$$D_{LS} = \frac{FC_{peak}}{VOLL} \qquad (1)$$

In (1), FC_{peak} is the investment cost (in \$/MW) of the marginal peaking plant, assumed to be the marginal provider of flexibility and reliability, and $VOLL$ is the value of lost load (in \$/MWh), that is, the VCR.

Interestingly, the optimal level of the reliability standard, in this case, is *independent of the shape of the load duration curve*. While such simplicity may be desirable in terms of computational complexity, this approach essentially *disregards risk*[18]. In contrast, as discussed later, consumers may be more generally seen as risk-averse, and so are, generally, network and system planners too. Furthermore, market risk might call for significant hedging costs on investors too, to the point of limiting investment. This aspect of risk is particularly important in a power system dominated by VRE and storage, as the load duration curve can change significantly year by year (again, see the discussions below).

For these reasons, **LOLP and LOLE may be unsuitable** as metrics to drive reliability standards, especially if used as *the only* metrics. **Different reliability metrics and forms should therefore best be**

[15] It should be noticed that the two are similar but not equivalent, which may bring confusion. In fact, for example an LOLP equal to 1 in 10 years means that a loss of load event is expected to happen *once* every ten years, which is different from 24 hours/10 years or 2.4 hours/year. In contrast, an *LOLE*=2.4 hours/year indicates that the expected total duration of interruptions in a year is equal to 2.4 hours.

[16] AEMC and Reliability Panel, "2022 Reliability Standards and Setting Review - Issues Paper", January 2022.

[17] S. Stoft, *Power system economics*, IEEE Wiley, 2002.

[18] There is an implicit *risk-neutrality* assumption in such a simple model of reliability, as also discussed by Stoft[17].

considered, and approaches that fully assess the amount of load and/or energy not supplied might be more suitable.

Energy adequacy, not only capacity adequacy

The historical focus of the planning exercise has been on ensuring that the installed capacity of available facilities could meet demand at *peak times*, with little (if any at all) regard for conditions outside the peak time(s). Hence, if we focus on generation, the planner's task has traditionally been to identify the right amount of *capacity or reserve margin* (defined as installed *conventional generation capacity* minus expected peak demand) that would guarantee certain predefined reliability standards. The issue has thus always been about generation *capacity* adequacy. However, with the uptake of energy-limited resources (ELR) such as different storage technologies and a more flexible demand side, the broader concept of **"*resource adequacy*",** including in particular *energy* adequacy, is more suitable to describe the reliability requirements of a low-carbon power system. Accordingly, **the way we *measure* adequacy might need to/should change too**. In particular, the concept of **capacity margin become obsolete and unsuitable**

Gap in the existing reliability framework in dealing with operational challenges of low-carbon grids

With deeper penetration of VRE and ELR, there is a pressing need to incorporate (more detailed) **operational studies and relevant considerations in planning studies** that drive reliability standards, particularly to deal with **emerging flexibility requirements that may lead to short-duration reliability risks**. Indeed, as demonstrated by the California rolling blackout event of August 2020, load disconnection might happen due to flexibility requirements and operating constraints[19]. This is evident in Figure 3, showing the significant ramp that conventional resources had to go through to cover the *net demand* in California during the August 2020 events. In particular, the figure highlights the presence of a still positive ramp between 5pm and 7pm, leading to the net demand peak at 7pm, while the underlying demand was already ramping down since 5pm (peak demand time).

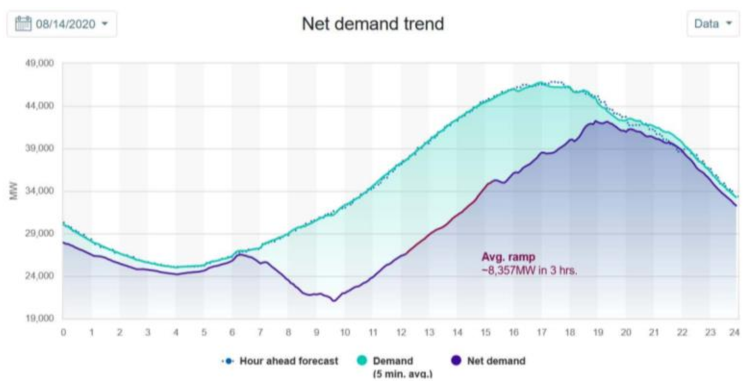


Figure 3. Hour-ahead forecast and real-time demand and net demand in California, 14 August 2020 (Source: Joskow's presentation at MIT, 5 October 2020, taken from California Independent System Operator[19])

[19] P. Joskow, "California's Blackouts, Near Blackouts, and Fires", MIT presentation, 5 October 2020.

Similarly, as witnessed in the NEM, high and extremely volatile prices may be largely due to lack of flexibility and ramping needs beyond what planners and investors would have envisaged. Such market volatility may also impact investments and thus adequacy in the long run. Based on the Rules[4], "*a secure operating state is one where the power system is in, or will return, to a satisfactory operating state following a credible disturbance*". In the examples just considered, net system ramps could indeed be taken as *credible disturbances*. However, at the same time, "*Clause 3.9.3C(c) of the NER specifies that a power system reliability incident is an incident that AEMO considers would have been avoided only if additional active energy had been available to the relevant region or regions from generation, demand response or inter-regional transmission elements*". Ramping may therefore be interpreted as a reliability issue too, in the sense that ramp scarcity might be resolved through additional capacity or energy being made available when required. Lack of flexibility is something that the planners and regulators should be concerned about, especially because of their impact on market prices and then resource investment. Based on the comments above, in order to send the right investment and operational signals to the market with regards to emerging flexibility and ramping issues, there is **a need for bringing closer operational (security) and planning (adequacy) considerations**. This becomes even more important while new markets for system services will influence investment in new technologies beyond the incentives provided by a reliability-constrained energy-only market.

Weather-driven disturbances and resource adequacy

While reliability has historically being associated to power plant scheduled or unscheduled outages as the main driver for risk, in future power systems, especially for planning purposes, the reference to **classical "discrete" generation contingencies per se will become less relevant** and the shift will be over to **identifying and forecasting weather patterns and relevant "disturbances" that will lead to new and key forms of contingencies/disturbances**. These will include disturbances **at different timescales**, including relatively short ones (**ramping** timescale of a few hours) as well as longer ones (**dunkelflaute** timescales, in the order of one to a few weeks).[20]

In fact, one of the most daunting tasks to ensure reliability in a low-carbon grid will be identifying the likelihood and features of *dunkelfaute* ("*dark doldrums*") events. Furthermore, in weather-driven systems, where both supply and an electrified demand are much more sensitive to weather variables, specific events that might be not extreme if taken individually might lead to an extreme impact due to the compounded effect on both supply and demand where these could dangerously move into "opposite directions" - for instance some type of heatwaves under calm winds causing higher demand and lower supply levels, or, as in Texas[21], extremely cold snaps with large part of the generation fleet in outage due to insufficient "weatherization".

However, **the existing framework does not address in anyway the challenges that dunkelflaute events will bring, especially in a system with large shares of ELR**. Furthermore, the range and variance of potential conditions and events that could lead to resource adequacy issues become much wider in a weather-fuelled system, raising the question of type and features of reliability standards to deal with such a volatile and diverse set. In this context, a reliability framework based on **LOLP/LOLE is**

[20] The assessment of these disturbances may be a challenging task due to data availability, especially when dealing with different datasets such as wind speed, solar radiance, temperatures, etc., compounded by the fact that it is likely that the future will look different from the past!
[21] Wikipedia, *2021 Texas power crisis*, https://en.wikipedia.org/wiki/2021_Texas_power_crisis.

again completely inadequate, in that it does not measure in its fullness the severity of such events, which could manifest themselves with very high impact in terms of energy not supplied, for example.

Figure 4, elaborated from Biggar and Hesamzadeh[22], shows the interplay of different reliability metrics through the load duration curve in a system with and without VRE, with renewable energy modelled as negative demand and assumed with no capacity value. It may be appreciated how, with reference to a generation system designed with reference to an LOLE reliability standard calculated from the VOLL and the fixed cost of the marginal producer, the introduction of VRE would have the effect of increasing significantly the size of the annual load shed while maintaining the same LOLE. The effect on the EENS would instead depend on the specific shape of the load duration curve. As it may be appreciated from Figure 5, following a simple traditional generation planning screening analysis approach, the optimal amount of capacity of each type of generation[23] in each scenario would also change, while the proportion of time each generator operates as the marginal generator would not. In practice, due to the variability of renewable output as well as potentially the uncertainty in the actual peak demand of the system, especially in the presence of flexible demand technologies, the net load duration curve would be (potentially very) different year by year, especially with regards to peak times. This would introduce significant operational volatility in the system and generation investment risk.

Although based on a very simplified model, this analysis illustrates the complexity of the interplay between reliability and economics based on traditional approaches based on LOLP or LOLE. Under the transitioning grid, other metrics such as the USE/ENS may be more suitable to drive the reliability standards design. In fact, in general, as an *integral* metric, the USE/ENS would provide information on *both* duration and magnitude of reliability events, as opposed to only probability or frequency of interruptions.

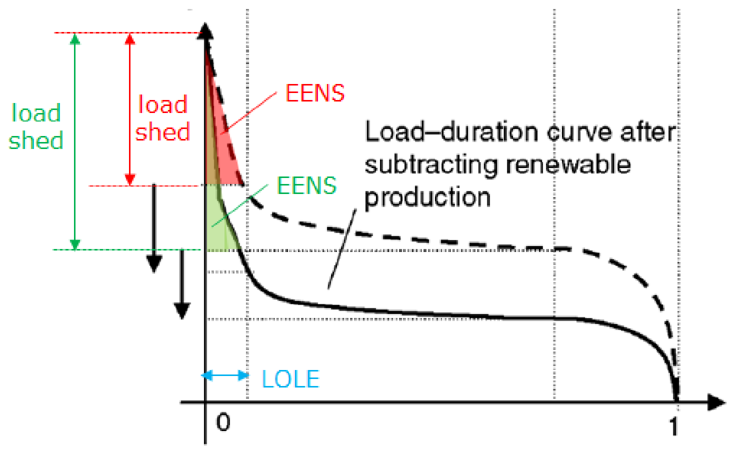


Figure 4. Interplay between reliability metrics and economics in a system with and without VRE (adapted from Biggar and Hesamzadeh[22]).

[22] R. Biggar and M. Hesamzadeh, *The economics of electricity market*, Wiley, 2014.
[23] This simple model considers a baseload generator with fixed cost f_1 and operational cost c_1 (the slope of the screening curve) and a peaking generator with fixed cost f_2 and operational cost c_2.

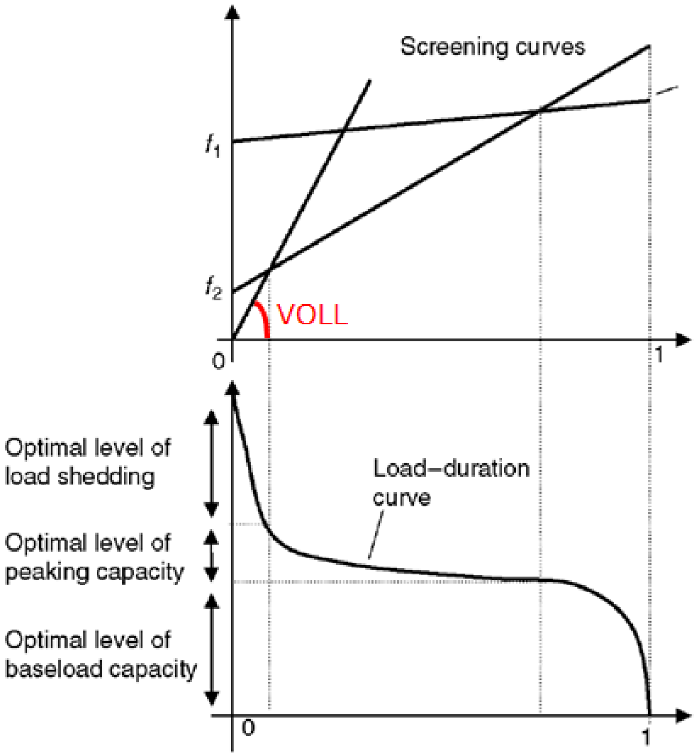


Figure 5. Illustration of classical screening curve analysis (adapted from Biggar and Hesamzadeh[22]).

Impact of storage in low-carbon grids

Storage will play an essential role in future system dominated by renewables. An illustrative example is provided in Figure 6[24], showing the trends of how storage, in this specific example in the form of pumped hydro storage plant (PHSP), of different power capacity (different rows of the table in the figure) and energy (different surface colours) levels, could crucially support provision of reliability in a system that becomes increasingly decarbonised and in which conventional generators retire.

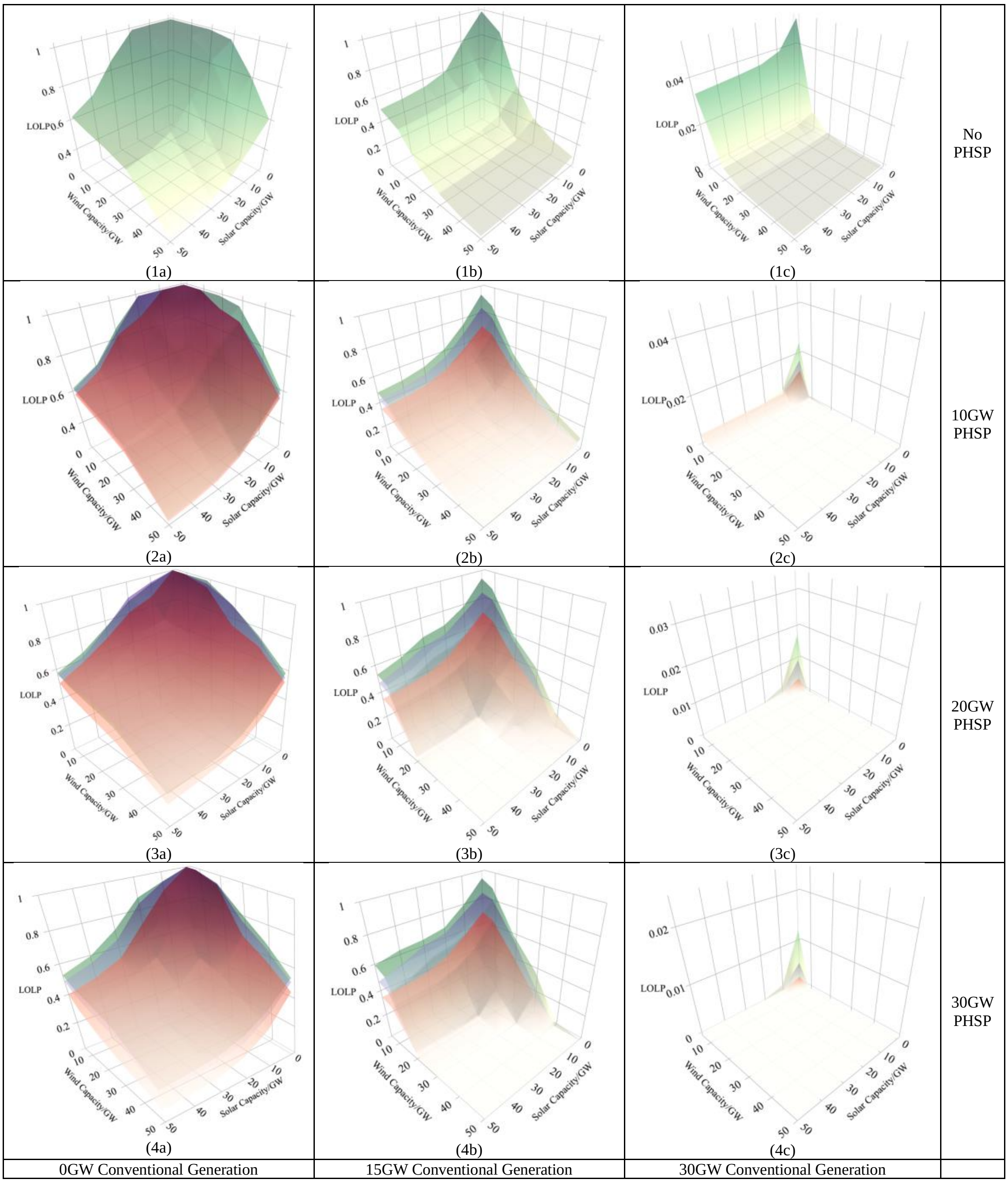


Figure 6. LOLP as a function of available conventional generation and PHSP storage. The three surfaces represent energy storage duration associated with power capacity: 6h (green),12h (blue) and 24h (red) (Source: Liu and Mancarella[24]).

[24] G. Liu and P. Mancarella, "Adequacy Assessment of Renewables-Dominated Power Systems with Large-Scale Energy Storage", Australasian Universities Power Engineering Conference (AUPEC) 2019, Fiji, November 2019.

However, its market behaviour may significantly affect the outcomes of its reliability contribution, and the relevant modelling is highly complex.

For example, Figure 7 illustrates some of the challenges that storage might introduce based on a case study taken from a PhD thesis at the University of Manchester[25]. The energy storage in question has power rating equal to 12% of the system peak (same for both charging and discharging) and energy capacity equal to 6% of the daily energy consumption.

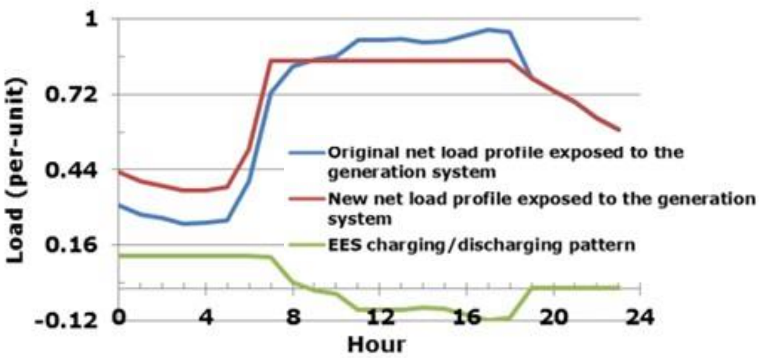


Figure 7. Illustration of the impact of storage on the change in the net load profile, based on a peak reduction storage control strategy (Source: adapted from Y. Zhou[25]).

As seen in Figure 7, by shifting on–peak energy consumption to off–peak, the impact of storage is to create a (much) longer and flatter peak. This chronological change in the shape of daily demand profile that needs to be the supplied by the generation system affects the resource adequacy outcomes. An illustration of the effects of this behaviour can be appreciated in Figure 8, showing the cumulative distribution function (red curve, vertical axis on the left-hand side) and probability density functions (blue histogram, vertical axis on the right-hand side) of the duration (in hours) of *individual capacity shortfall events* for the two cases with (left) and without (right) storage.

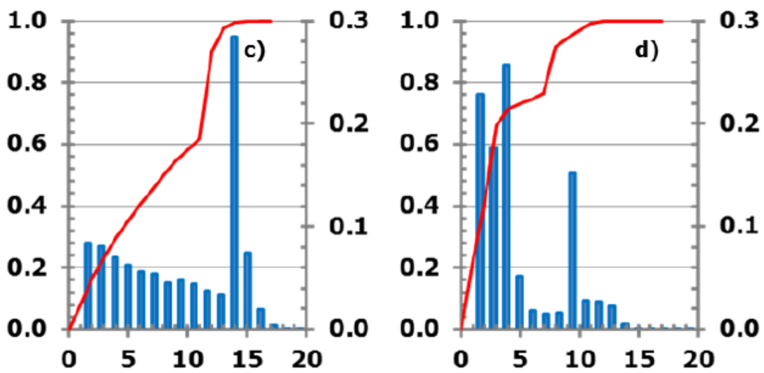


Figure 8. Probability distributions of the duration (in hours) of individual capacity shortfall events with storage (left) and without storage (right). The red curves represent the cumulative probability distribution functions (vertical axis on the left-hand side), and the blue histograms are the probability density functions (vertical axis on the right-hand side). (Source: adapted from Y. Zhou[25]).

[25] Y. Zhou, *Comprehensive framework for assessment of the contribution of demand response and electrical energy storage to power system adequacy of supply*, PhD Dissertation, The University of Manchester, 2015.

The overall impact is that while storage decreases the magnitude of the peak, it also increases significantly its length, exposing it for longer time to the risk of the supply side to fail. This is translated into a change in the risk profile of interruption events, whose distribution of duration becomes wider and more volatile, including much longer duration cases (left picture) than in the base case without storage (right picture). While this example is only illustrative, this behaviour is rather typical for control strategies that aim at reducing the peak demand.

Furthermore, it is likely that **the interaction between renewables and energy storage**, including in hybrid plants, **will lead to even higher degrees of volatility and further change the risk profile when measured through different metrics.** In this regard, Figure 9, taken from an internal, illustrative analysis conducted by the author's research team, shows how the distribution of the Energy Not Supplied metric changes for increasing energy levels of storage (columns from left to right) and increasingly decarbonized scenarios (rows from top to bottom). Different storage power capacity levels are also parametrized in each figure through different colours. From visual inspection, the illustrative results suggest that with deeper penetration of renewables the risk profile shifts towards higher impact events. Eventually, shallower storage options might not be able to provide suitable solutions to decrease such a risk, and **a storage portfolio mix with both large power capacity and long duration solutions might be needed while we decarbonize the NEM**. Overall, these considerations again point to the need for **adopting both a set of different metrics as well as tail indicators to fully capture future system risk**.

Risk-aversion in reliability decisions

Risk-based approaches are based on considering both the *probability* of occurrence of different scenarios *and* their potential impact, particularly to describe the wide range and volatility of relevant potential events. Building resource adequacy against extreme events might be particularly challenging. However, as demonstrated by the 2021 events in California and Texas, it is exactly in those stress-tests that a reliability framework should demonstrate its robustness. For example, De Vries and Jimenez[26] discuss a new methodology (European Resource Adequacy Assessment - ERAA) that has recently been adopted in Europe, which requires that resource adequacy assessment methodologies should consider different scenarios with an estimate of their likelihood, including scenarios with extreme weather driven for example by climate change. It should be noted, in this regard, that occurrences of extreme events somehow associated with inadequate planning might be less tolerated than extreme events due to operational contingencies, as the perception of the public and the market stakeholders might be that there has been some form of negligence from policy makers, regulators, etc., the moment that better planning could have avoided such events (see more below). At the same time, with more electrified sectors, the impact of interruptions may become more and more significant and potentially dramatic.

On these premises, the rule maker might also want to consider ***risk aversion*** when building some form of resilience to extreme events in the context of reliability standards. However, expected values of reliability metrics, as used so far, are more reflective of a **risk-neutral** approach, in which more extreme cases would be washed out within the general averaging of all events, with no actual hedging against extreme volatility and impact. For example, in theory the EENS reliability performance of a system that had experienced a very large event like the 2021 Texas blackout might still be within the average

[26] De Vries and Jimenez, "Market signals as adequacy indicators for future flexible power systems", *Oxford Open Energy*, 2022, 1, 1-5.

standards set for a certain period! Even worse, LOLP and LOLE metrics would not even be able to capture the true impact and implications of such events! Hence, no effective adaptation plan could be put in place, unless the metrics were changed.

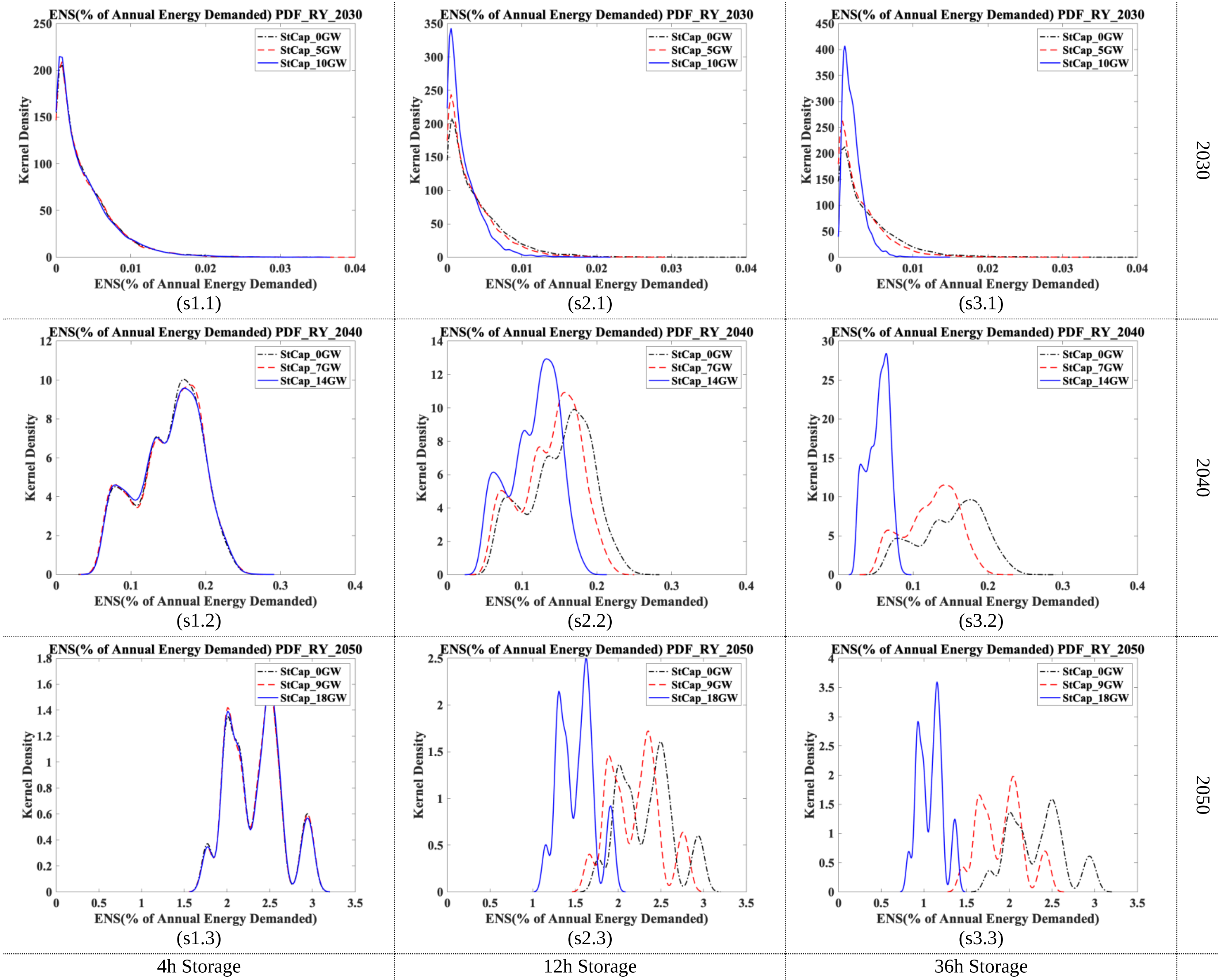


Figure 9. Distribution of ENS under Different VRE Penetration and Storage Capacity Levels (Source: author's research team)

Socio-economic rationale for risk-aversion

As from the above discussions, decision-making approaches to designing reliability standards have, more or less implicitly, assumed risk neutrality, whether based on engineering judgement or economic theory, using a combination of LOLP/EENS and the value of lost load. However, in reality, the public may become increasingly reluctant to accept interruptions of supply, especially in rich countries, and this “social” driver should be considered along with techno-economic factors in the design of new standards. The fact itself that many customers and communities consider more and more local backup options, and even “going off grid”, is a clear sign of lack of tolerance to any interruptions, especially major ones. It is difficult to think differently after events like the Texas one that even led to loss of lives. The Australian Energy Regulator (AER) had indeed initiated a discussion in this direction via the “WALDO” project[27]. Also, once again, while there might be some degree of tolerance for operational security issues given the challenges of low-carbon system operation, the public and various stakeholders might be less lenient to interruptions due to resource adequacy.

Fairness and cost allocation: who benefit from and who pays for reliability?

From a customer perspective, reliability standards should drive the system and market towards socially optimal solutions. However, is making “average” decisions fair? Eventually, when extreme events happen, it is most likely that disconnection would affect more significantly customers without backup options. On the other hand, tail-based settings could further push the demand side to be active and, in case, price-responsive, thus unlocking new reliability providers and revealing the true utility functions and risk attitude, including of flexible customers that might want to trade some degree of reliability for economic benefits. Overall, inclusion of tail metrics and adoption of risk-averse approaches should also be able to provide fairer outcomes across the whole range of network users.

[27] Australian Energy Regulator, “Widespread and Long Duration Outages - Values of Customer Reliability - Consultation paper”, March 2020.

5. Measuring reliability in low-carbon grids

Reliability performance of the new system

The much higher volatility we can expect in system and market operation in the context of deep penetration of variable energy and energy-limited resources will completely change the risk profile of low-carbon grids. While the type of metrics[28] adopted for reliability assessment insofar appears to be generally suitable, given that the relevant probability distributions of these metrics also become much more volatile, **multiple metrics might be needed that provide more information about the actual system's risk**, again especially for risk-aversion purposes**.** This is for instance shown in Figure 10, taken from the latest ESIG report on resource adequacy[29], which illustrates tail risk and further suggests how arrangements such as based on LOLP/LOLE may no longer be fit for purpose in a world dominated by VRE and ELR.

Overall, a desirable feature of a new reliability framework is that the system be capable to deal effectively with both expected and less expected events, possibly **limiting the number, frequency and magnitude of interruptions, especially in the context of risk-averse users of the system**. **Specifications as to what reliability performance might be expected in the presence of relatively extreme events** may also be helpful, for example to limit the maximum duration or power loss in the case of an extreme event.

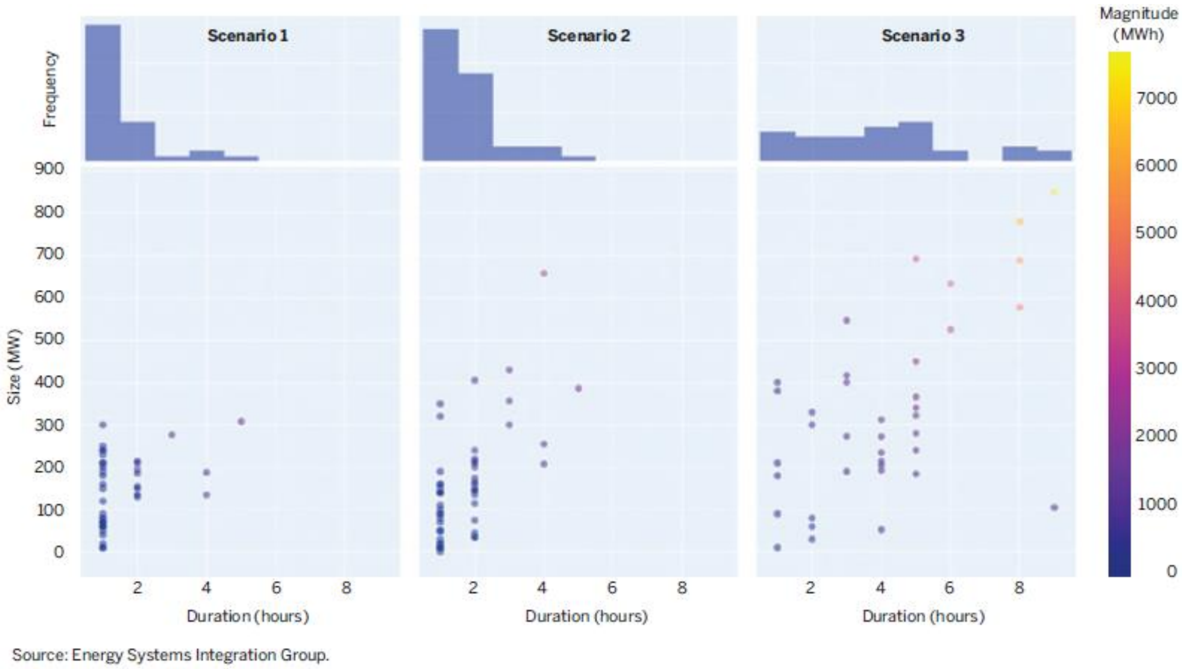


Figure 10. Scatter plot of size, frequency, and duration of shortfall events with increasing reliance on energy-limited resources (ESIG, 2021[29])

[28] For example, Expected Energy Not Served (EENS), Loss of Load Probability (LOLP), Loss of Load Expectation (LOLE), etc.
[29] ESIG, "Redefining Resource Adequacy for Modern Power Systems", 2021.

Need for considering the distribution tail

For a given reliability metric such as unserved energy, the expected value captures the entire probability distribution. However, it crucially misses any emphasis on the tail, which is where lie the most extreme situations resource adequacy should provide for. Moreno *et al.* have made these considerations in several papers[30].

Brito-Pereira *et al.* [31] illustrate a few case study examples that draw attention to the tails, especially looking at the potential stresses that climate change could cause in terms of frequency and impact of extreme weather events. Some examples are actually taken from AEMO, with regards to their forecast of events exceeding the existing USE reliability standard in New South Wales (Figure 11)[32], while others refer to the Colombian electricity market, assessing the impact on prices coming from long duration water scarcity events prompted by El Nino (Figure 12)[31]. As the Colombian system is 70% reliant on hydropower, this may be a good representation of what the future could look like in system more dominated by ELR in the presence of scarcity event. The market volatility driven by physical scarcity is evident, and it is interesting to note that not even the 99th percentile of the price distribution is fully capturing the tail! **More studies should be performed on energy-limited hydro-dominated power systems and the suitable mitigation options that have been put in place in countries such as Colombia or Brazil.**

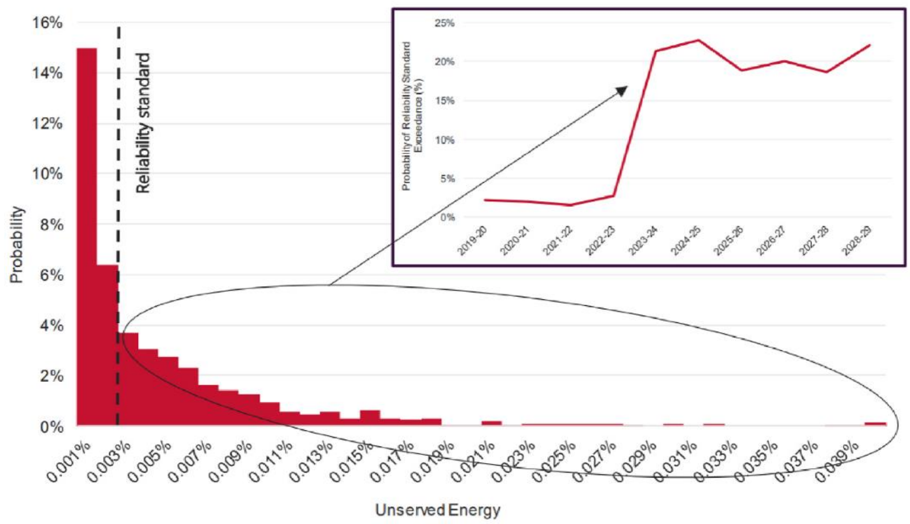


Figure 11. Probability distribution of annual USE in New South Wales, 2023–2024 (AEMO, 2019[32]; figure taken from Brito-Pereira *et al.*[31]).

[30] See for example: R. Moreno *et al.* "From Reliability to Resilience: Planning the Grid Against the Extremes", *IEEE Power and Energy Magazine*, July-August 2020.
[31] B. Prito-Pereira *et al.*, "Adjusting the aim of capacity mechanisms: Future-proof reliability metrics and firm supply calculations", *Energy Policy*, 2022.
[32] Australian Energy Market Operator, "2019 Electricity Statement of Opportunities", 2019.

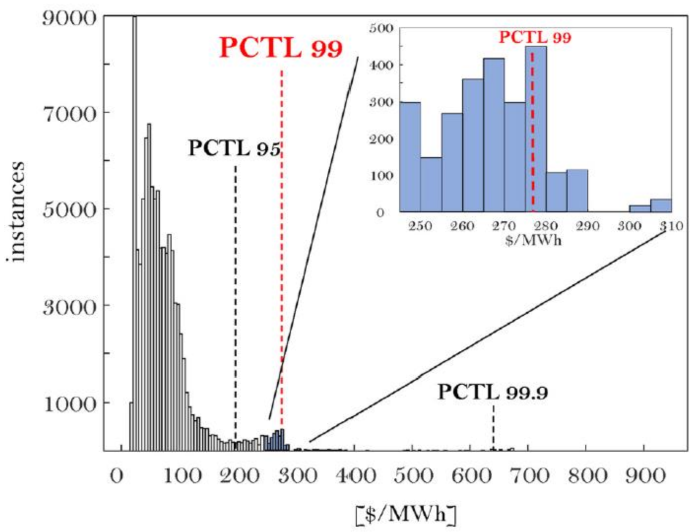


Figure 12. Probability distribution of wholesale electricity market prices in Colombia between 2009-2018, to illustrate scarcity events (Brito-Pereira *et al.*[31]).

In fact, an indicator that has been recently proposed to better assess the *"fatness" of the tail* is the "Conditional Value at Risk" (CVaR). This is defined as the mean value of a preselected percentile of worst cases (the "tail") in the probability density function of a relevant indicator, for instance, the ENS (Figure 13)[30]. For example, an α%CVaR represents the expected unserved energy in the worst (1-α)% cases. A 99%CVaR would thus correspond to the expected energy not supplied in the worst 1% cases.

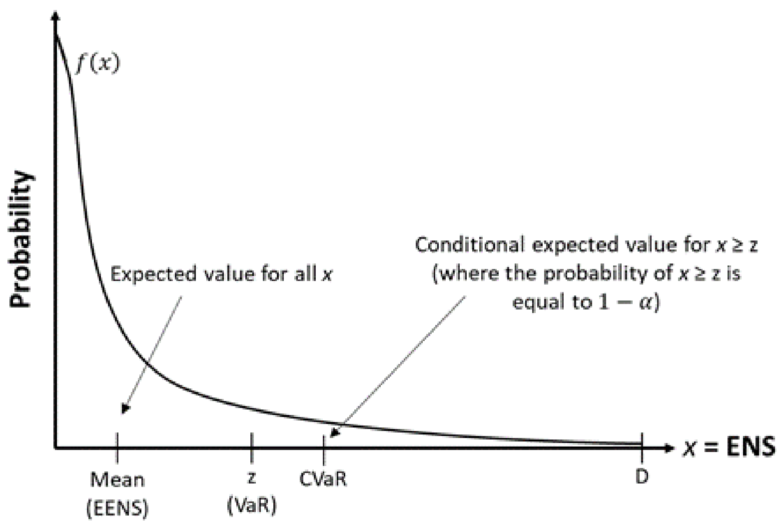


Figure 13. Probability distribution of energy not supplied, highlighting the "mean" metric (EENS) and "tail" metric (CVaR), adapted from Moreno *et al*[30]. 1-α indicates the size of the considered set of worst cases, *VaR* (or *z*) is Value at Risk, *CVaR* is Conditional Value at Risk, *D* is maximum demand, *x* is Energy Not Supplied (ENS), and the Mean value is the Expected Energy Not Supplied (EENS).

<u>*Why risk-aversion may be desirable in spite of potential over-investment*</u>

It is possible that risk-averse standards would lead to over-investment relative to risk-neutral positions, which however in the long term would likely pay back through risk hedge in terms of providing more stability to the market. In this respect, Stoft[17] highlights how even in a conventional system, which is much less volatile than a low-carbon system, year by year volatility associated with scarcity prices can pose significant stress on investors. Within an energy-only market, and in the context of a risk-neutral world, there is already the well-known tension between relying on volatility and scarcity pricing to drive operating and investment incentives and the fact that too much volatility may eventually be undesirable for investment. However, moving forward, in a very low-carbon grid a key question will be as to what extent **an increasing challenging risk profile (e.g., manifesting itself as higher and higher intra- and inter-annual volatility) might start curbing investments to the point that a risk-averse perspective would/should eventually emerge for the market to be able to deliver reliability of supply**.

Reliability, price volatility and demand side flexibility

As argued by De Vries and Jimenez[26], in a non-far future, storage and flexible demand, also enabled by sector coupling and digitalization, might mean that at times of high VRE output these price-responsive technologies could change the traditionally inelastic demand curve by introducing different slopes. Conversely, under supply scarcity, flexible demand could reduce their consumption depending on its willingness to pay and storage might even start producing based on their assessment of the incumbent opportunity cost. Hence, this demand-side flexibility might mean that no interruptions would occur, but supply scarcity would manifest itself through high prices. Such behaviour has already been witnessed in different instances in the NEM, with price spikes also needed to allow cost recovery for thermal power plants operating at lower capacity factors and to create sufficient arbitrage opportunities for storage plants. While from a purely economic perspective this might not seem as a problem and could potentially even be regarded as a sign the market is working properly and provide effective signals, issues such as intertemporal linkages in ELR (which are a large component of flexible demand resources), ramping scarcity[33], and so forth, might mean that in an energy-only market excessive reliance of demand side flexibility might no longer effectively signal reliability risk and eventually curb investment on the supply side. Furthermore, given the variability of possible situations that might arise, there could be times when energy limits could become binding in flexible demand resources too, eventually also leading to interruptions of "inflexible" customers.

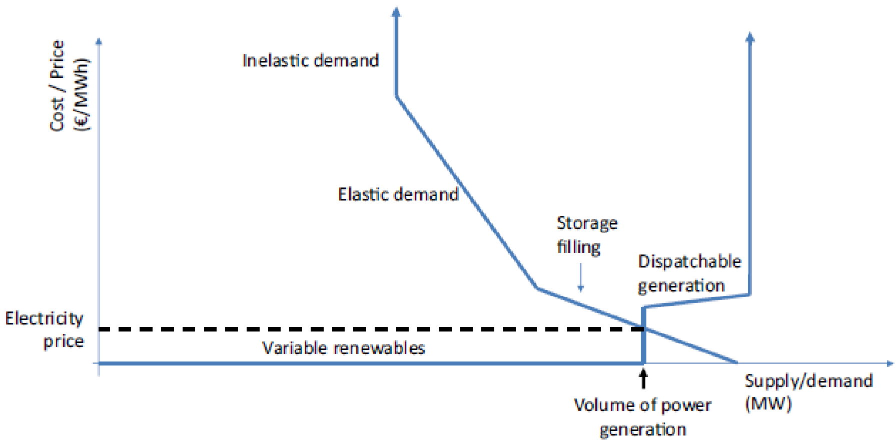


Figure 14. Effect of storage and demand side flexibility on electricity prices with high VRE output (Source: De Vries and Jimenez[26]).

[33] Due to intertemporal constraints, ramp scarcity may manifest through high/low price signals outside the actual times when there is a need for ramping.

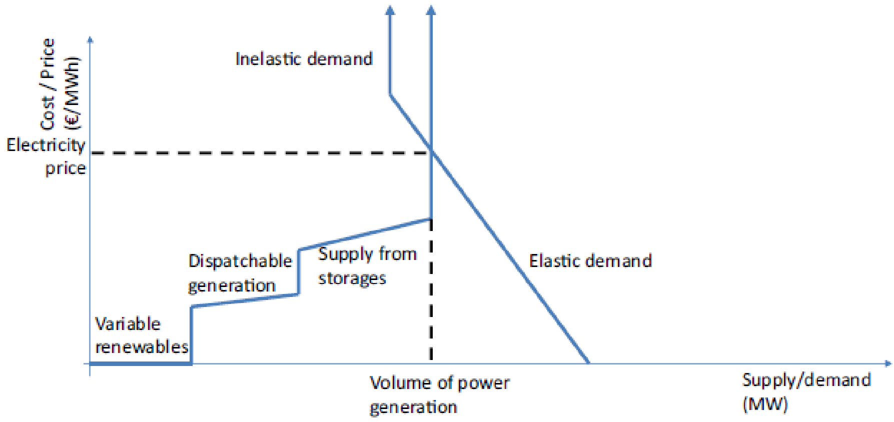


Figure 15. Effect of storage and demand side flexibility on electricity prices with low VRE (Source: De Vries and Jimenez[26]).

General economic theory about demand response argues that it might significantly reduce market volatility. However, given the very diverse characteristics of available resources and uncertainty that may be expected in the supply side and operation of distributed energy systems, in practice low-carbon grids could witness potentially significant *volatile prices* regardless and in spite of the increasing demand side flexibility. High prices for excessively long periods or occurring very frequently should eventually signal inadequate investments even without interruptions – which might not take place thanks to demand flexibility[34]. However, **the current approach to reliability standards would not capture periods of flexibility scarcity**, since, as we discussed above, the standards would be driven by either engineering considerations or by a simplified model of reliability that assumes inflexible demand (and therefore that resource inadequacy would manifest itself through load disconnection).

Furthermore, price volatility might signal new requirements for the system that have not been traditionally captured in resource adequacy studies. For example, as mentioned earlier, lack of flexibility would correspond to price spikes, and price volatility in general is a sign of system inflexibility[35] - as again already seen in the NEM. As argued by Joskow[36], high prices might be as socially unacceptable as interruptions (especially with electricity playing a more prominent role in heating/cooling and transport too) and might eventually call for regulatory interventions that could be avoided in first place by suitable designing of the reliability standards. Excessive price volatility should also be avoided to reduce investment risk, as discussed above[37].

On the other hand, with demand becoming more elastic, there might be a portion of flexible demand that would not be supplied simply because the value it attributes to electricity is lower than the market clearing price. As discussed by Prito-Pereira *et al.*[31], a possible idea is that such demand should not be considered as unmet in the calculation of the reliability standard, with the result that the assessment,

[34] I. Perez-Arriaga, *et al*. "*Utility of the Future: An MIT Energy Initiative Response to an Industry in Transition"*, 2016. https://energy.mit.edu/wp-content/uploads/2016/12/Utility-of-the-Future-Full-Report.pdf.
[35] J. Cochran, *et al.* "Flexibility in 21st century power systems". *Technical report. National Renewable Energy Lab. (NREL),* 2014. https://www.nrel.gov/docs/fy14osti/61721.pdf.
[36] P.L. Joskow, "From hierarchies to markets and partially back again in electricity?: responding to decarbonization and security of supply goals", pp. 1–17. https://doi.org/10.1017/S1744137421000400.
[37] Contracts could also be used for risk hedging to a certain extent, but there will be a point where excessive volatility should suggest that there is a fundamental resource scarcity issue.

and even definition, of the "actual" unserved energy becomes more difficult to carry out (Figure 16). They further suggest that in practice high-price events might be included into reliability assessment by setting *price thresholds*. When market prices were to rise above these thresholds, loss of load metrics could be augmented accounting for these hours as if there were an outage, while unserved energy metrics could be augmented by the energy volumes cleared above the thresholds, to identify flexible demand components. Insofar, only the Belgian system operator, Elia, seems to have *considered* this idea of price thresholds for augmented reliability standards, particularly for capacity market mechanisms and resource capacity value assessment. Eventually, Elia's final implementation was performed by including "near-scarcity" hours, that is, times when a marginal demand rise *would have led* to load disconnection[38].

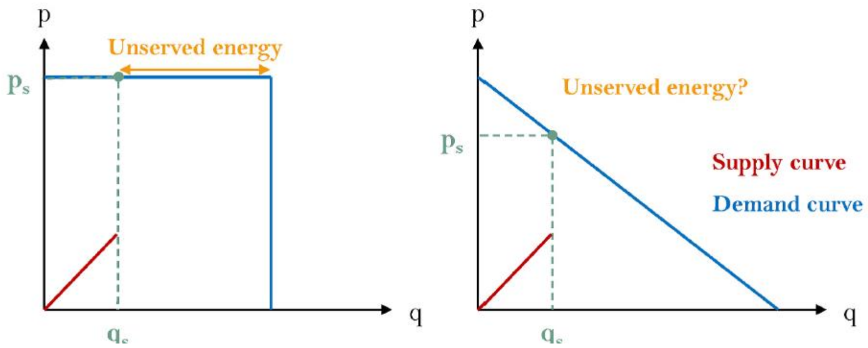


Figure 16. Unserved energy with inelastic (left) and elastic (right) demand: with very flexible demand, the assessment and even the definition of unserved energy become more difficult (Source: B. Prito-Pereira *et al.* [31]).

[38] *Elia*, "Overview of Belgian CRM Design: Introduction Note", 2019.

6. General considerations and recommendations for a new, evolved reliability framework

Based on the different discussions carried out above, this section attempts to summarise the key findings and considerations and propose a way forward for the development of an evolved reliability framework that could best account for physical and economic risks that might emerge in low-carbon power systems.

Capturing weather variability and associated inter- and intra-annual volatility

Given the huge variability of weather across years, it is essential that reliability standards be able to **capture the full spectrum of possible events** that might arise in order to build the most adequate portfolio of resources.

In particular, in the context of a weather-driven system dominated by VRE and ELR, **a standard/reliability framework should clearly address short-term operational risks (e.g., from ramping) as well as FOR-driven and energy limit-driven events that could cause serious stress for longer duration weather events, including extreme events (e.g., "dunkelflaute")**.

Explicitly considering risk aversion

Under the emerging techno-economic conditions of an evolving grid, **traditional (often implicit) assumptions of risk neutrality of consumers, planners, regulators, and even investors may no longer hold**. Therefore, **reliability metrics should allow for risk-aversion considerations too**. Hence, in the context of an extended definition of reliability that should **also include more extreme events**, a reliable system should remain secure and adequate to all plausible disturbances. Of course, there might be a need for indicating a set of extreme conditions to deal with, as well as a "degree" of *resilience* that the system would achieve by limiting its degraded state following an extreme event and by featuring a prompt recovery.

The question therefore arises as to how account for risk-aversion in a more volatile supply and demand environment. In this sense, while reliability has traditionally employed expected values of metrics with reference to measuring the system performance under generally normal, expected conditions (and credible events), as important elements of a new reliability framework **more extreme conditions should be *explicitly* measured, assessed and possibly limited via consistent or even the same type of metrics used for expected value analysis**.

Modelling the full distribution of events

Analysis of the probability density functions (PDFs) of the traditional metrics (ideally more than one) used to assess reliability could provide a way forward that **could inform development of reliability standards to fully account for the risk of a more volatile weather-based system, presence of extreme events and energy-limited resources, and risk-aversion of different agents**.

This practically corresponds to **considering in different ways "tail" values of the PDF of the metrics that measure the system's reliability performance, and not only their "expected" values**.

The reverse question also arises, that is, as to whether focusing on extreme situations and creating reliability standards that account for tail conditions would also satisfactorily address the more "normal" range of expected situations that may arise. In this context, in principle a system that were to be to some extent "resilient" to more extreme and volatile conditions, *might not* be reliable, in the sense that it may not meet minimum "classical" reliability standards, for example based on annual average measures of EENS, LOLE, etc., which would have been primarily designed having in mind more expected situations.

Therefore, **expected value and tail metrics should be *jointly* adopted to better describe the full probability distribution of impacts and thus actual reliability risk**.

Considering more than one metric for robust decision making

There may also be rationale for adopting more than one metric alongside expected and tail values for a given metric.

This is fundamentally based on ensuring that, given the extreme variability and volatility of potential physical outcomes, **reliability standards should be able to capture all the features of future events**. For example, very frequent but small events could also be tolerated much less in the future. These considerations point to the **need for limiting both integral metrics *as well as* the impact of individual events (in size, for rare events, and/or frequency, for smaller events),** especially if there were a significant risk of unmitigated extreme impact[39].

Identifying the exact boundaries of the events to be included in the analysis should carefully be considered in modelling exercises to inform the development of suitable standards. It is thus also recommended that **a suitable modelling/assessment roadmap that could effectively address the requirements for new type of studies and modelling be developed** (see more below).

Options for risk-aware reliability standards

As discussed, from a mathematical perspective **risk aversion** can be built by using the *tail* (rather than the expected value) of the probability distributions that are used for reliability assessment[40]. This can be obtained by **introducing tail risk measures such as the Conditional Value at Risk (CVaR)**[41] **applied to the distributions of the energy not supplied, loss of load, or other relevant metrics**[42].

Reliability standards could then be described as a (weighted) combinations of *average* – and therefore risk-neutral – probabilistic measures (e.g., EENS, i.e., the expected value of the ENS probability distribution function) **and *tail* – and therefore risk-averse – probabilistic measures** (e.g., the CVaR of the ENS probability distribution function)[43].

For example, a **composite reliability metric** R_{ENS} for energy not supplied could be set up as follows:

[39] Expected value modelling could also be applied to the tail itself, and *event-specific* integral metric could be integrated over all relevant extreme events, to be compared and measured with the likes of metrics such as EENS. This is essentially what a CVaR approach does.

[40] See for example: G. Strbac, D. Kirschen and R. Moreno, "Reliability standards for the operation and planning of future electricity networks", *Foundations and Trends in Electric Energy Systems*, Vol. 1, Issue 3, 2016.

[41] As illustrated in Figure 13, taking the ENS metric as example, an α%CVaR essentially represents the expected energy not supplied in the higher (1-α)% cases. A 99%CVaR would thus correspond to the expected energy not supplied in the worst 1% cases, while the "classical" EENS metric would represent the expected energy not supplied over all 100% cases.

[42] Other options, particularly suitable for planning, are to adopt methodologies such as robust optimization or minmax regret that specifically hedge against the occurrence of worst-case scenarios. See for example: R. Moreno, A. Street, J. M. Arroyo, P. Mancarella, "Planning low-carbon electricity systems under uncertainty considering operational flexibility and smart grid technologies", *Philosophical Transactions Royal Society A*, Vol. 375, Issue 2100, Aug 2017, pp. 1-29.

[43] There are different ways of formally approaching the issue in the context of mathematical programming, in general by using CVaR in either the objective function or the constraints of the relevant optimization problem. The specific design of the reliability standards might be informed by such techniques – see also below.

$$R_{ENS} = w \cdot EENS + (1 - w) \cdot \alpha\%CVaR_{ENS} \qquad (2)$$

Furthermore, by changing the relative weights assigned to expected value and CVaR components, risk-aversion could also be further modulated and thus controlled, moving from a fully risk-neutral approach (w=1) to a fully risk-averse one (w=0), thus leading to a fully *risk-aware* decision-making model (see further below).

A similar approach could also be applied to other reliability metrics, and, if multiple metrics were to be adopted, a portfolio of different metrics and expected and tail measures could be optimally designed to hedge against the estimated risk profile.

Cost-benefit analysis to determine risk-aware reliability standards and their settings

If **risk-aversion** needs to be considered in reliability standards to inform market developments and build the opportunity to hedge against volatile scenarios, a core consideration/challenges is as to how to **develop a methodology to perform an appropriate cost-benefit analysis (CBA)** to determine the numerical value of the standard settings.

The methodology would require risk-averse models such as again based on CVaR. For example, instead of determining the optimal energy not supplied level as a tradeoff between cost of interruptions, assessed based on the USE (basically, the EENS, that is, a risk-neutral measure) and investment into the marginal reliability provider level, the CVaR (e.g., at 95% level) of the distribution of the energy not supplied could be used to augment the EENS, as illustrated in (2). Further elaborations on possible CBA approaches are provided in Section 7.

Other similar emerging decision-making options could be adopted too. For instance, one could consider the use of ***least-worst weighted regret (LWWR)*** approaches, such as the ones being adopting by National Grid in the UK (and recently AEMO too) for transmission planning[44]. A *regret analysis* would replace the more traditional cost-benefit analysis currently adopted and based on expected cost and expected unserved energy across scenarios. While a LWR approach has traditionally be looked at as fully risk-averse, its recently proposed weighted counterpart, the LWWR, could provide more options for risk attitude modulation and thus risk control[45]. Sensitivity around the weights could also provide a powerful tool to assess the robustness of the decisions made, which is the way National Grid is practically deploying the methodology.

Finally, for the case of multiple metrics, a multi-criteria methodology could be devised that would assess optimal trade-offs among costs, energy not supplied, frequency and duration of events, etc., and again both for expected as well as tail realizations.

Flexibility scarcity and price volatility

Besides being interpreted as early signs of system stress and technical inadequacy for specific features (not only capacity, but also flexibility, as discussed above), even in an environment with significant responsive demand, **price volatility and excessively high/low prices ("near scarcity" prices) could** thus also **be incorporated into new reliability standards**. This would also provide **a further tool to**

[44] P. Mancarella, *et al.*, "Study of advanced modelling for network planning under uncertainty", *Report prepared for National Grid Electricity System Operator*, 2020. "Part 1: Review of frameworks and industrial practices for decision-making in transmission network planning", https://www.nationalgrideso.com/document/185821/download; "Part 2: Review of power transfer capability assessment and investment flexibility in transmission network planning" https://www.nationalgrideso.com/document/185826/download.

regulators to identify scarcity events.

Of course, modelling flexibility scarcity and capturing it at 5-minute resolution will be challenging given data input and computational requirements, but it should be at least considered. For instance, **arrangements that are a development of the current forecast uncertainty measure (FUM) approach**, with augmentation that would consider price volatility and threshold metrics, could be a possible way forward for suitable operationalization of new standards.

On the other hand, some might simply argue that, as new entrants are generally more flexible than those leaving the market, the problem might not arise. However, even if this assumption were correct, it might lead to reliance on a service that it is not explicitly compensated for or incentivised.

General summary recommendations

In line with the above considerations, it is recommended that **elements that should be part of revised reliability standards** for low-carbon grids dominated by VRE and ELR could include:

- **A set of more than one metric, with preference for integral metrics that address aspects of energy not supplied, so that different features of both expected and more extreme events could be fully captured**, particularly with regards to inter- and intra-annual renewable energy output variability;
- **Information on the full probability distributions of the metrics selected to describe the system performance**, for instance via augmenting expected value measures with "tail" indicators;
- **Risk-aware approaches that could account for a modulated risk attitude of decision makers, for example based on CVaR** or alternative CBA methodologies such as based on least-worst weighted regret (LWWR) analysis;
- **Operational issues such as coming from flexibility scarcity and ramping requirements;**
- **Near-scarcity events, particularly in the presence of more responsive demand.**

7. Considerations on the design of a composite reliability metric

Identifying efficient levels of overall risk

A first key problem to address in introducing risk-aversion within reliability standards is how to identify efficient levels of overall risk that could/should be accepted, including tail and average basis risk.

A potential approach that could be explored is based on **budget-constrained incremental investment for tail reliability.** This would rely on **setting up a certain additional budget** (e.g., 20% of the budget allowed for average value-driven reliability investment) **to be allocated to investment for tail reliability** (extreme scenarios) relative to the investment required to meet expected values of reliability. A possible methodology is outlined as follows.

The **"classical" system reliability problem** aims at minimising investment cost *INV* plus operational cost *OP* plus the cost $C(E\{x\})$ of expected levels of (un)reliability $E\{x\}$ (or, similarly the expected reliability cost, depending on the specific formulation and indices used) associated with a certain metric *x* (e.g., the energy not supplied, whose expected value is multiplied by the value of lost load to obtain the reliability cost of interest):

$$\min INV + OP + C(E\{x\}) \quad (3)$$

For practical applications such as to identify suitable reliability metrics, standards, and settings, and to guarantee certain minimum levels of reliability to customers, once the value attached to reliability is known (e.g., the VOLL), the problem may be recast by constraining the operation and investment cost by a suitable level of desired reliability[45]:

$$\min INV + OP \quad (4)$$

$$s.t. \quad E\{x\} \leq \epsilon \quad (5)$$

Where ε is the relevant reliability standard setting for the expected value of the reliability metric *x*.

Two similar options can then be considered to **augment the above problem and include tail metrics**:

$$\min INV + OP \quad (6)$$

$$s.t. \quad E\{x\} \leq \epsilon \quad (7)$$

$$CVaR\{x\} \leq \bar{\varepsilon} \quad (8)$$

where $\bar{\varepsilon}$ is the relevant reliability standard setting for tail metric, or alternatively:

$$\min INV + OP \quad (9)$$

$$s.t. \quad w \cdot E\{x\} + (1 - w) \cdot CVaR\{x\} \leq \tilde{\varepsilon} \quad (10)$$

where $\tilde{\varepsilon}$ is relevant reliability standard setting for the composite average-plus-tail metric.

[45] See Strbac *et al.*, 2016, where it is also highlighted that reliability can be constrained locally or regionally to ensure certain *fairness* levels to customers. Furthermore, the objective function $\min INV + OP + C(E\{x\})$ could also be directly used to explore optimal level of the setting ϵ, to be used in the formulation (4)-(5). This is effectively the approach used to draw the efficient reliability level curves discussed above.

Let us focus on the formulation (6)-(8) as it allows separate and more explicit consideration for average and tail risk[46]. The key problem here is to determine the tail setting value $\bar{\varepsilon}$ once the value for the "traditional" expected value setting ϵ has been determined (e.g., through the procedures already employed in the current reliability framework).

A possible approach to determine the optimal tail setting is based on the following formulation:

$$\min\ \bar{\varepsilon} \quad (11)$$

$$s.t.\ INV + OP \leq B \quad (12)$$

$$B = B^0 + \varDelta B \quad (13)$$

where B is the total allowed reliability "budget", broken down into budget to ensure the "base", expected value level of reliability, for reliability B^0, and $\varDelta B$, which is the ***incremental*** **budget set aside to deal with tail risk**. This incremental budget is, essentially, the **"willingness to pay"** for tail reliability and thus **to create an insurance allowance against high-impact events** whose probability is very difficult (if even possible) to estimate, and could be determined according to different considerations, including socio-economic and political ones, as further discussed below.

This formulation yields an optimal (minimum) level of tail risk to be considered given a certain budget and once the "base" level of reliability and the relevant budget B^0 have been found based on (4)-(5). This is also exemplified in Figure 17, illustrating how starting from the expected reliability level ϵ and its budget B^0, and once allocated an incremental tail risk budget $\varDelta B$, it is possible to extract the optimal tail risk setting $\bar{\varepsilon}$ (indicated as starred to highlight that it is an optimal value).

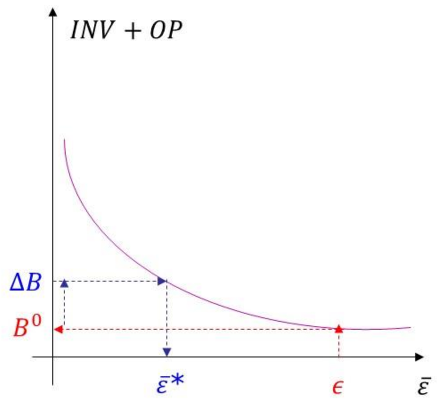


Figure 17. Illustration of the determination of tail reliability metric setting starting from the expected reliability metric setting and relevant budget as well as budget allowance to deal with tail risk

Such an approach, basically **adopting a parametric/sensitivity analysis to study the implication of different tail metric setting values, has the benefit that it can explicitly account for a limit determined by a certain budget that may be available**. What this budget actually is may depend, as mentioned earlier, on different considerations, including socio-economic and political one, e.g., how averse a population or a political environment might be to supply disruption. A reasonable upper limit

[46] The formulation (9)-(10), which makes explicit use of the composite reliability standard formulation (2), can also be reconciled with the formulation (6)-(8).

for this budget could be determined by the value of lost load times the expected value of the reliability metric in the tail, that is, the CVaR.

Furthermore, and again going back to the formulation (6)-(8), this approach also has the advantage that the *shadow costs* for both expected and tail risk could be determined and studied jointly, for example to assess their impact on other reliability settings used for market operation such as the market price cap (MPC) and cumulative price threshold (CPT).

Parameters of a composite reliability metric

Another key challenge in modelling risk aversion is to effectively represent the risk appetite of the relevant individual decision makers (being they planners, regulators, investors, generic market stakeholders, etc.). In the context of the form of reliability standards and relevant settings, this translates into how to determine the relevant parameters of a standard or associated methodologies, e.g., how to select the weight *w* as well as the *α-level* of the CVaR for the composite reliability metric in (2), repeated here for clarity:

$$R_{ENS} = w \cdot EENS + (1 - w) \cdot \alpha\%CVaR_{ENS}$$

- ***CVaR α-level***

With regards to the *α-level*, its selection should be based on both the **perceived appetite for risk** in a given socio-economic and political context as well as **analysis of the underlying techno-economic factors** that may determine the risk profile. For example, a relatively rich country that had recently experienced blackouts of some type would likely naturally set on higher values to cover a most extreme part of the tail and thus the potential impact on consumers that would be normally expecting an extremely high reliability level regardless of the underlying disruption causes. On the other hand, technical studies to determine the risk profile of the specific problem under consideration could also be helpful. For example, reliability studies aimed at determining the probability distribution of the unserved energy (or other selected metric) could highlight the likely presence of a more or less heavy tail under specific scenarios, as for instance in the case of the studies performed by IES for their RSSR modelling report, and highlighting the presence of heavy tails in unserved energy in both base case and low renewable output scenarios[47].

- ***Weight w in composite reliability metric***

As for the weight *w*, consideration for the most suitable split between average and tail events could be informed by a **parametric analysis**.

For example, taking certain reliability metrics and measures (in this case, expected and tail unserved energy), objectives (e.g., least-cost investment that guarantees minimum standard levels), and decision variables (e.g., a set of resources that are predefined investment candidates), studies could be run for the **boundary cases *w*=0** (tail value only, **full risk aversion**) and ***w*=1** (expected value only, **risk neutrality/no risk consideration**) as well as, possibly, **a number of (1-*w*) intermediate steps**, with the aim of building the associated Pareto-efficiency frontier. The decision maker could then take an

[47] IES, *2022 Reliability Standards and Settings Review – Modelling report*, Fig. 52.

informed decisions based on different considerations, including the incremental economic and risk benefits/impacts across the two dimensions from moving from one solution to another.

An example of such approach is shown in Figure 18, taken from Lagos *et al.*[48], and showing the impact of different investment solutions, including a sketch of the Pareto front, on the expected energy not supplied (EENS) and the *conditional* expected energy not supplied (basically, the CVaR) in the case of earthquakes.

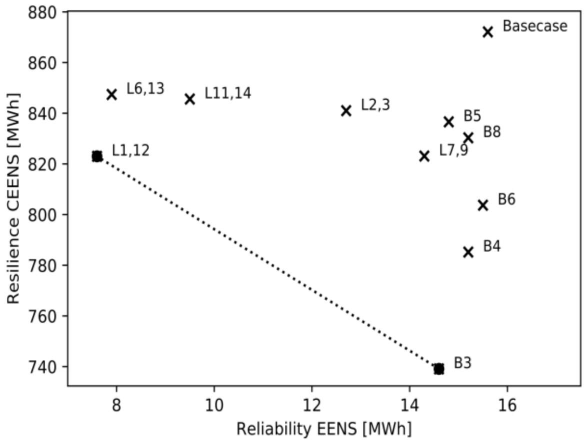


Figure 18. Example of Pareto front built across expected (reliability) and tail (resilience) risk metrics (Source: Lagos *et al.*, 2020[50]).

In practice, once selected the CVaR *α-level*, the same studies currently performed to assess the **impact of different investment options** against the expected unserved energy could be repeated **with focus on the tail events/scenarios appearing in the CVaR**. An illustration of this approach is shown in Figure 19, showing the efficient reliability level curves for different technologies, i.e., how the cost (e.g., annual investment plus reliability cost) of different new entrant investment options (indicated with coloured crosses) could be assessed for the EENS (as done in the current framework for the USE studies) and the CVaR of the energy not supplied. A comparison between the results from the two sets of studies could thus bring insights into costs and benefits of different technology solutions from a multi-criteria perspective and how investment options traditionally designed on the basis of expected values could perform in terms of more extreme events (and the other way round).

A key benefit of such an approach is that it allows **a clear assessment of the role of different technologies** in meeting specific objectives, and particularly **to deal with rare events**. For instance, such analysis could highlight whether specific technologies might be particularly suitable or unsuitable in dealing with extreme scenarios of interest, and contrast their performance under extreme cases against their performance under expected cases. This could possibly also **inform the incremental investment requirements to uplift the average reliability performance to account for high impact low probability events** that was discussed above. Based on the discussions throughout this report,

[48] T. Lagos, *et al.*, "Identifying Optimal Portfolios of Resilient Network Investments Against Natural Hazards, With Applications to Earthquakes", *IEEE Transactions on Power Systems*, Vol 35, Issue 2, pp 1411 - 1421, Mar 2020.

this kind of stress-test for different technologies is expected to be of growing importance in future systems dominated by renewables and energy-limited resources.

A similar analysis could then also be performed parametrically by **changing the value of the weight *w* and, in case, of the *α-level* too, for the composite reliability metric** (Figure 20), in order to, for instance, **assess the implications of changing the risk aversion** (the weight) for overall cost and reliability levels, as well as for the role to be played by different technologies.

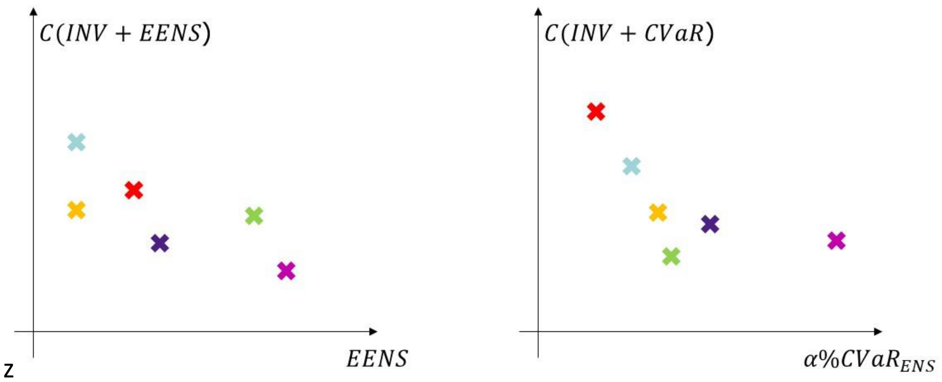


Figure 19. Illustration of efficient reliability level plots for (left) expected energy not supplied (*w*=1) and (right) *α%* conditional value at risk of the energy not supplied (*w*=0).

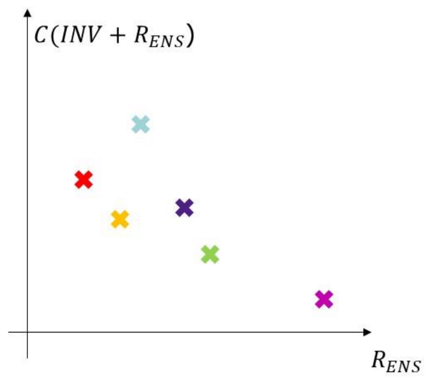


Figure 20. Illustration of efficient reliability level plots for a composite reliability metric $R_{ENS} = w \cdot EENS + (1-w) \cdot \alpha\%CVaR_{ENS}$, for a given level of the weight $0 \leq w \leq 1$.

As aforementioned, the outcomes of the above studies would eventually provide an optimal set of technologies associated with relevant costs and reliability performance, measured through different metrics (e.g., EENS and $CVaR_{ENS}$ or a composite metric with a given weight). The decision maker could then compare the different results to study the cost and reliability implications of how different types of technologies could meet specific objectives, including how they would perform under expected and extreme conditions, and to further inform potential incremental investment analysis as in the methodology outlined earlier for optimal risk level identification. **Key value of such studies, eventually, would lie in the insights that could be gained through *comparative analyses* of different scenarios and solutions, even more than the specific outputs per se**. On the other hand, deriving different efficient reliability level curves would require a significant number of studies based on computationally demanding simulations (see also below). However, this is something that would be advisable to carry out at least in the initial developments leading to the setup of new reliability metrics, especially to explore the potential implications of different arrangements and settings and gain more insights into the role of different technologies.

It should also be noted that parametric studies of this type might also be most useful to **inform the numerical settings of the composite reliability metric** as an alternative to or to complement the budget-constrained approach discussed above. For example, **the cost implications of different new entrants could be explored against a certain selected composite metric** (say, for the sake of example, a metric that would tend to be risk-averse, e.g., with 95%CVaR and w=0.8), exactly as it is currently done for an average value metric. Alternatively, once again **efficient reliability level curves could be separately drawn against expected value metric and tail metric** (Figure 19), and the results reconciled *ex-post* based on different kinds of considerations. This approach could also be suitable to determine different types of settings associated with the reliability standards, e.g., the market price cap (MPC) and the cumulative price threshold (CPT) across a predefined time window (e.g., one week or two weeks).

Market settings associated with a composite reliability metric

In the formulation (6)-(8) the dual variables of the two reliability constraints could be used to inform relevant pricing mechanisms associated with settings such as the MPC and CPT. Similar outcomes could be obtained through the parametric studies discussed earlier.

Regardless of the specific methodology used, one possible idea to explore could be to develop ***two sets* of MPC/CPT pairs**, namely, one for average reliability, e.g., (MPC_{EENS},CPT_{EENS}), and the other for tail reliability, e.g., (MPC_{CVaR},CPT_{CvaR}).

The two different MPC levels could be associated with the different investments (in potentially different technologies) required for the marginal entrant to provide expected and tail reliability and should be determined alongside the two CPT values. Furthermore, the length of the window within which the CPT values should be assessed could also be driven by the specific risk profile emerging from technical and techno-economic studies. For example, "expected" events could be associated with CPT determined within a one-week window, while "tail" events – particularly if concerned, in a renewables-dominated world, about *dunkelflaulte* events – could be linked to a two-week CPT period, if this were to be supported by the technical findings about the drivers for tail risk.

How to move from one level of settings to the other would of course be a challenge and requires careful considerations and studies. One idea could be to **introduce a triggering point** for the shift, which could be dictated by both market performance and technical (e.g., weather) forecast, with for example a step change from one to the other as opposed to a gradual shift (e.g., increasing linearly the value of the parameters).

Further challenges for a market environment

While the whole process to determine reliability and associated market settings for a composite reliability metric is already a very difficult task per se, in a market environment, and specifically an energy-only one, further challenges exist that might hinder the applicability of different approaches.

For example, while the formulations and methodologies discussed above may be suitable to drive the design of the reliability settings, in practice the outcomes might be that, in order to recover investments aimed at hedging against rare but high impact events in the tail, **very high prices close to**

the value of lost load and for a relatively long and sustained period of time should be allowed. Whether this would be socially acceptable and practically implementable deserves extensive discussions.

Furthermore, in a market context, where decision making is decentralised, an additional challenge lies in how (and whether at all) different risk appetites from different stakeholders could be reconciled. For example, the system operator or regulator might be naturally more risk averse than market participants in proposing and setting up market operating rules and incentives. However, given the rarity of tail events, which in principle might not materialise in many years, **market stakeholders might naturally be inhibited from investing into technologies whose return would principally rely on very volatile occurrences**. In other words, if market stakeholders were to value their projects based on *expected*, risk-neutral returns, the resulting investment decisions might not align with signals emerging from a risk-averse methodology.

One possible approach to mitigate the distress for investors in resources aimed at providing tail reliability could be, in designing the trade-off between MPC and CPT, to go, given a certain budget, for a relatively higher MPC and shorter CPT. In this way, less-rare events could be captured too while enabling technology revenue adequacy in a "safer" way. Of course, the downside of this approach is that the long duration reliability requirements signals, for example for deep storage, might be diluted too and even distorted, which might defeat the purpose of the whole design if not done carefully[49].

Eventually, **rare events and risk-aversion call for certain types of insurance policies, whose mechanisms and pricing may be fundamentally different from traditional decision making that is driven by expected cost and benefits.** There is also a general perception that eventually it is a Government's responsibility to make sure that suitable measures are taken to contrast the impact of extreme, even if foreseeable, events. In this context, as discussed in the Cigre Working Group C4.47 "Power system resilience", one could argue that **markets alone might *not* be able to deliver on the (operational and investment) requirements for extreme events**[50]. This justifies the **rationale of different types of interventions** that could be put forward to specifically deal with extreme events, such as a development of the Reliability and Emergency Reserve Trader (RERT) mechanism, different types of strategic reserves, government-backed investments and incentives, etc[51].

What is essential to stress, however, is that such **insurance policy mechanisms should be efficiently coordinated with market operation and price signals**, for example following the guideline principles outlined here, and that the presence of such mechanisms to deal with events of unique nature such as high impact low probability ones should and ideally would **enhance the scope and performance of a market,** *and not replace it*. Moving forward, insurance-like strategic reserve mechanisms could also

[49] Similar challenges might be associated with other risk-averse methodologies such as LWWR if decentralised decision-making approaches do not align with the methodology to design incentive schemes and there may be more suitable for more centralised schemes, such as for network planning or for capacity payments/markets where the procurement level were to be determined centrally.

[50] P. Mancarella *et al.*, "Economic and regulatory aspects of power system resilience", Working Paper for the Cigre Working Group C4.47 "Power system resilience", 2022.

[51] Specific capacity mechanisms to deal with tail events might also be designed, although of course there would be other market design issues associated with them, particularly to efficiently integrate the design of the energy and capacity markets and for both expected and extreme events. Different types of capacity mechanisms, including more decentralised schemes, could also be devised.

be designed that explicitly take into account the differentiated reliability level of different consumers and the growing role of flexible demand side resources[52], as also discussed above.

Modelling framework to robustly model evolving tail risk

From a techno-economic modelling perspective, several challenges are associated with the requirements for studies able to draw a full risk profile of future systems and inform on new reliability requirements, standards and settings.

First of all, and as already discussed earlier in this document, care should be taken in determining the scenarios of interest, particularly to identify the drivers that concurrently could lead to the occurrence of rare events of high impact. In future systems, this calls for integration of **more detailed weather models**, for both the supply and demand sides and relevant technologies, into reliability studies.

Probabilistic studies should then be adopted, for example based on extensive Monte Carlo simulations, and again it is essential that the underlying random variables are extracted from probability distributions that are adequately designed to include rare events. Computationally, then, the ability to sample tail events in a Monte Carlo framework is a well-known challenge *per se*, and suitable techniques could be considered to improve computational efficiency and allow an adequate number of studies to be run[53].

New features should also be included into the modelling that might not have been traditionally considered in reliability studies for resource adequacy, particularly, as again extensively discussed above, to **capture ramping and operational flexibility requirements**.

Another aspect that might be important to capture is to study and then **simulate more realistically the behaviour of energy-limited resources**, particularly longer duration ones, in the presence of extreme events that lead to the activation of relevant MPC and CPT settings, as it may not be straightforward to identify their opportunity cost, and therefore their market-driven behaviour, in the presence of extreme operational (price- and weather-driven) uncertainty.

[52] See for instance: F. Billimoria *et al.*, "An insurance mechanism for electricity reliability differentiation under deep decarbonization", *Applied Energy*, 2022.

[53] See for example: G. Liu, *et al.*, "Assessment of the capacity credit of renewables and storage in multi-area power systems", *IEEE Transactions on Power Systems*, 2021".

8. Concluding remarks and next steps

This briefing note, prepared for the Australian Energy Regulator Market Commission (AEMC) with regards to the 2022 Reliability Standards and Settings Review (RSSR) and considering their relevant papers published in 2022, has outlined the main drivers for change for the development of an evolved reliability framework. A number of technical and economic considerations, based on fundamental comparisons between historical and low-carbon power systems, have been made as the basis for a number of recommendations that have been provided, including new approaches to better assess the full risk profile of an evolving system and potentially determine new standards forms.

As next steps, the suitability of some of the recommendations made should be checked against potential practical applications, particularly with regards to including a full characterization of the risk profile of future systems via updated metrics and to considerations for a new decision-making framework to inform the form and settings of the new metrics. Finally, a roadmap outlining the requirements for the procedures and modelling that would be needed to capture the new system's risk profile and inform an evolved reliability framework should be considered.

Acknowledgments

The author would like to gratefully acknowledge all the fruitful discussions with and the contributions and inputs provided by Dr Rodrigo Moreno, University of Chile.

Contact: Prof. Pierluigi Mancarella, pierluigi.mancarella@unimelb.edu.au