Level 22
530 Collins Street
Melbourne VIC 3000

Postal address
GPO Box 2008
Melbourne VIC 3001

T 1300 858 724
F 03 9609 8010
E info@aemo.com.au

5 May 2023

Charles Popple
AEMC Commissioner and Chair of Reliability Panel
Australian Energy Market Commission
Sydney South NSW 1235
Submitted online
REL0086

Dear Mr Popple,

Issues Paper – Review of the Form of the Reliability Standard and Administered Price Cap (APC)

This letter and attachment constitute AEMO's submission to the Issues Paper, published 30 March 2023, for the Review of the Form of the Reliability Standard and Administered Price Cap ('the review').

AEMO appreciates the opportunity to provide comment on the Issues Paper and actively participate throughout the review process. AEMO supports the Panel undertaking this review and agrees with the Panel's assessment of the need to account for tail risk in the reliability standard.

Feedback is provided in the attachment. This feedback should be considered as AEMO's initial consideration and will be further developed through engagement with the Panel. Feedback is provided across:

- Issues with current reliability framework
- Assessment of reliability metrics
- Modelling approach
- Reliability Metrics
- Administered Price Cap.

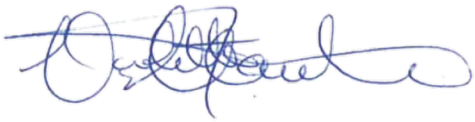
In summary:

- AEMO understands the theoretical basis of increased variable renewable energy (VRE) leading to increased tail risk, and agrees that the introduction of a tail risk measure appropriately mitigates this contingency.
- AEMO considers reliability should be communicated in terms of tangible measures for consumers such as depth, duration, and frequency.
- AEMO's preferred metric for consideration is a multi metric standard that is expressed as: the average annual outcomes that are at or above a 1-in-10-year probability may not be greater than x% of average regional load shed for 4 hours (or equivalent). This is statistically similar to the use of the conditional value at risk (CVaR) as a measure of tail risk, but expressed more tangibly for customer outcomes.

AEMO also notes throughout its submission that these reliability measures are restricted to the current National Electricity Rules (NER) 3.9.3C definition of unserved energy (USE), which considers only supply shortfalls arising from single credible contingencies on generators, and inter-regional transmission elements. This excludes any consideration for single credible contingencies on the intra-regional transmission networks, and all outages arising from multiple or non-credible contingency events. As most real-world outages fall outside this defined framework, it is important to acknowledge through this review that the customer experience of power system reliability is likely to be worse than forecast.

If you have any questions please contact Kevin Ly, GM Reform Development & Insights, at Kevin.Ly@aemo.com.au.

Yours sincerely,



Violette Mouchaileh
Executive General Manager – Reform Delivery

Contents

Summary

This submission explains AEMO's initial consideration of the Panel's review of the form of the reliability standard and APC.

The Panel should note the distribution of forecast reliability events indicates large events can occur with reasonable regularity despite the forecast also indicating low annual average expected USE. The current form of the reliability standard, aiming for an annual average expected USE of 0.002%, leaves the National Electricity Market (NEM) exposed to reliability events of a size that AEMO considers intolerable to consumers and policy-makers. This may lead to jurisdictional bodies seeking to achieve a higher level of reliability than the NEM provides. Rather than have this occur, AEMO welcomes the Panel's desire to adjust the form of the reliability standard to account, in some way at least, for the distribution of forecast reliability events so the market better manages large 'tail' events on an ongoing basis from 2028.

It must be recognised that the current form of the standard, annual average expected USE, is superior to other metrics, like Loss of Load Expectation (LOLE) and Loss of Load Probability (LOLP), because, by being a measure of energy directly, it accounts for size of the reliability event as well as probability. Being an average value, it naturally includes the large events in the tail of the distribution of simulations. The single annual average expected USE also allows direct calculation of the Reliability Settings, to stimulate the market to invest and provide for the standard. Further, annual average expected USE recognises that a level of reliability cannot be guaranteed, no matter how desirable that would be.

The drawback with the annual average expected USE is the 'averaging out' of the expected USE, such that it does not describe the type of reliability event consumers can expect and is therefore not tangible to consumers or policy-makers. These stakeholders, rather than understand what their average supply feels like, would prefer to understand the depth, duration, and frequency of reliability events that they may reasonably expect.

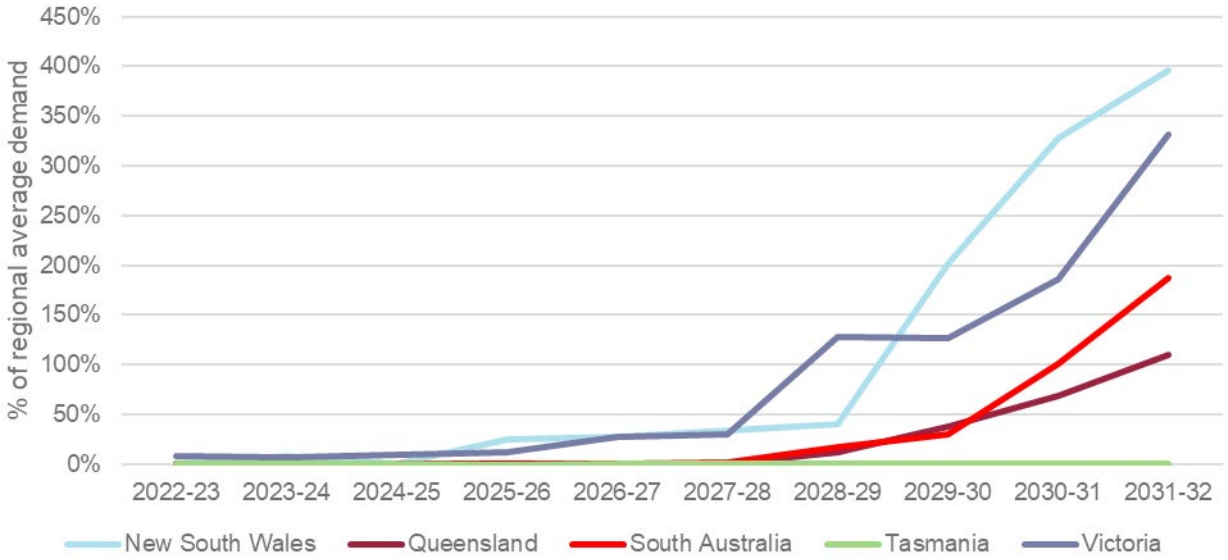
The Panel should take the opportunity to consider simpler or more tangible options than the annual average expected USE, which attempt to describe the depth, duration, and frequency of reliability events that consumers can reasonably expect. This is opposed to more complex mathematical approaches that may add further layers of complexity to the annual average expected USE that consumers and policy-makers presently struggle to engage with. An additional way to make the Standard more tangible is to add a predetermined condition, such as high demand, low renewables output, or both. However, unless these conditions are the factors that directly influence the distribution of forecast reliability events, they may detract from the standard, or if set onerously, may require a level of reliability that is uneconomic.

It is for these reasons this submission puts forward a preferred **multi metric standard**, based on **depth**, **duration**, and **frequency**, but is also calibrated and can be expressed to a value of annual average expected USE. The submission discusses the equivalence of expressing average USE in a standard based on depth and duration and proposes AEMO's preferred metric as the average annual outcomes that are at or above a 1-in-10-year probability may not be greater than x% of average regional load shed for 4 hours, or equivalent. This metric is statistically similar to the use of 10% conditional value at risk (CVaR) but expressed more tangibly for consumers.

This provides some tangible description of reliability events that consumers can reasonably expect, recognises that a level of reliability cannot be guaranteed, and retains the benefits of the annual average expected USE, that provides for the Panel to be able to calculate the appropriate Reliability Settings '. A

sample of the outcomes of this metric is shown below, based on data from the *Update to the 2022 Electricity Statement of Opportunities*.

Figure 1 The average depth of average regional load shed for 4 hours, or equivalent, for annual outcomes that are at or above a 1-in-10-year outcome



This submission also discusses the Administered Price Cap and supports the Panel investigating the options put forward in the Issues Paper.

Issues with the reliability framework

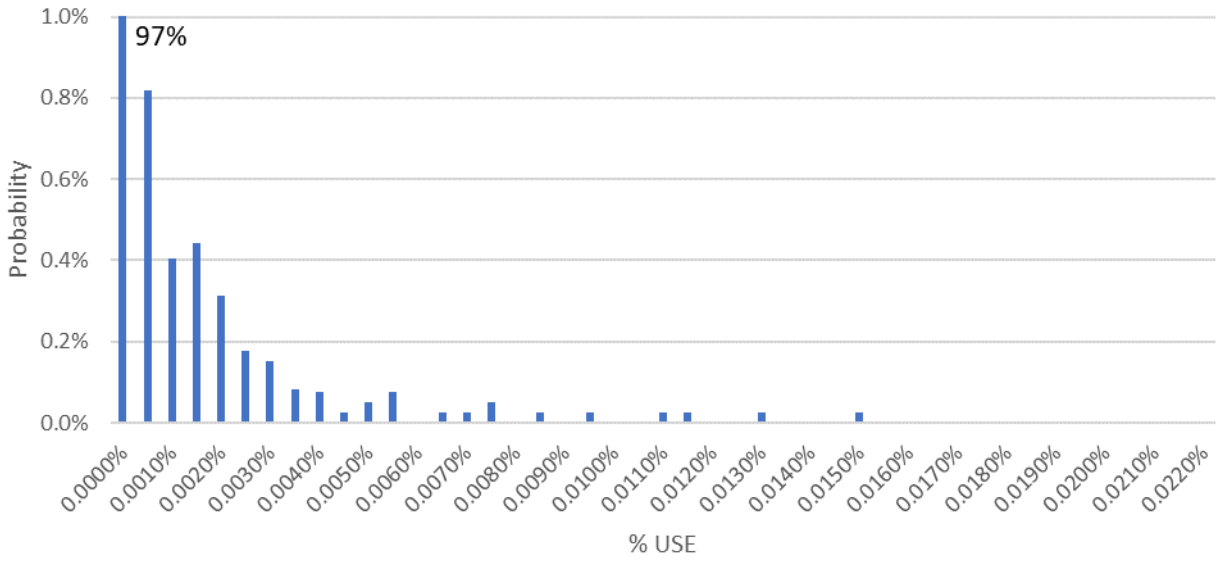
The changing nature of reliability events must be accounted for in the standard

AEMO supports the Issues Paper's characterisation that reliability needs to be thought about differently as the NEM transitions from a capacity limited to an increasingly energy limited power system. The Issues Paper requests feedback on the way reliability risks with the current framework are set out. AEMO agrees that the current reliability framework does not align with future reliability risks, understands the theoretical basis for increased VRE leading to increased tail risk, and agrees that adoption of a tail risk measure mitigates the risk of these events.

As part of the 2022 Reliability Standards and Settings Review (2022 RSSR), AEMO supported the Panel and technical advisor Dr Pierluigi Mancarella's characterisation of the changing reliability risk and AEMO considers that these recommendations have been effectively carried over to the Issues Paper. AEMO welcomes the analysis and identification of the changing distribution of USE outcomes and agrees that the range of considerations described on page 18 of the Issues Paper will require further development throughout the review process.

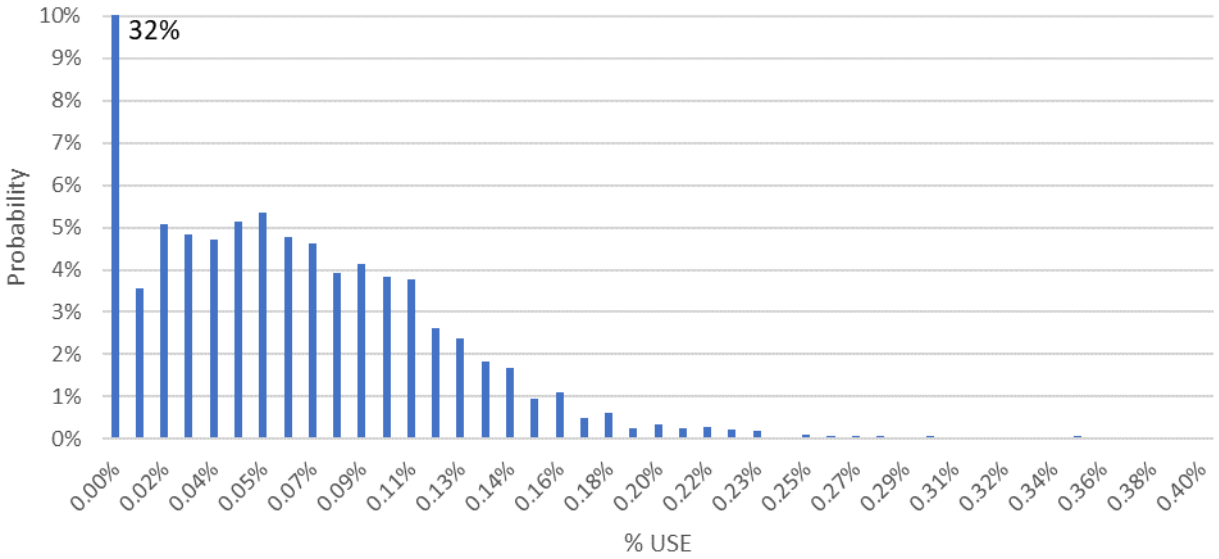
Figure 2 and Figure 3 below show the probability density functions of forecast average annual USE for New South Wales in financial year 2023 (FY23) and FY32 based on the February 2023 *Update to 2022 Electricity Statement of Opportunities*[1].

Figure 2 Probability density of forecast USE – NSW FY23



[1] https://aemo.com.au/-/media/files/electricity/nem/planning_and_forecasting/nem_esoo/2023/february-2023-update-to-the-2022-esoo.pdf

Figure 3 Probability density of forecast USE – NSW FY32



The figures show that in FY23, USE outcomes are skewed towards a smaller percentage USE event, with the x-axis showing small buckets of forecast USE. While these figures demonstrate a changing skew of the distribution of USE, given the large difference in the mean of USE – 0.00008% USE in FY23 compared to 0.055% USE in FY32 – it is difficult to determine whether there is a fundamental shift in the shape of the distribution based on the Electricity Statement of Opportunities (ESOO) methodology.

Modelling using alternative assumptions (including assumptions of additional VRE build) as described in the Panel's proposed modelling approach may better identify the nature of the changing tail risk profile. This supports the Panel's proposal to model distributions of reliability events, at certain target levels of annual average USE, such as 0.002% USE.

Customer experience of USE is worse than forecast

Section 4.3 of the Issues Paper defines 'tail risk' for the purposes of the reliability standard, as the events at the extreme end of a non-normal probability distribution. AEMO considers it important to recognise, however, that the operational realities and customer experience of power system reliability are likely to be worse than forecast, and broader than USE outcomes driven by defined reliability events.

The NER 3.9.3C definition and modelled distribution of USE are limited by the drivers of reliability events which does not include security events, multiple and non-credible contingencies and outages of network elements that do not significantly impact the inter-regional transmission capability.

AEMO considers this definition to be is extremely limiting, and while this definition is retained, the customer experience of power system reliability is not reflective of what the reliability standard provides for. AEMO notes that most outage events are not classed as reliability incidents due to the restrictions of this definition.

As the generation fleet becomes more geographically diverse, and the power system becomes more interconnected, this definition is likely to widen the difference between the forecast and the true consumer impact of power system reliability.

For example, the following events would likely not be classified as a reliability incident to the degree of the specific unavailability, and are therefore not considered as a possibility in reliability forecasts:

- Any supply shortfall that occurs coincident with a transmission outage that occurred due to a single credible contingency event on an intra-regional flow path (such as Dederang-South Morang).
- Any supply shortfall that occurs coincident with a prolonged double-circuit outage that occurred following a multiple contingency event on a key inter-regional flow path (such as the Heywood Interconnector).
- Any supply unavailability that occurs coincident with a prolonged outage of numerous generating units that occurred following a non-credible contingency event on shared connection assets.

AEMO considers that any load shedding that arises from the above circumstances would likely be characterised by consumers and jurisdictions as a power system reliability incident, despite their specific NER3.9.3C exclusion. Such stakeholders might also reasonably assume that power system reliability frameworks plan for and manage such events.

Further, USE modelling is based on a calculated level of reliability that may be technically possible based on optimised participant behaviour, but that could not be expected to occur. Deviation from fully optimised behaviour to manage portfolios, particularly for energy limited generation, will result in USE outcomes that are not considered within reliability forecasts. This will most likely increase the risk of USE above modelled probability distributions, particularly as generating units respond to operational circumstances not included in the definition of USE.

AEMO acknowledges that while the definition of USE and the defined modelling requirements are not directly in scope of this review, it is important to note and consider these limitations when developing the tail risk profile and considering the form of the reliability standard. Most of the power system reliability outcomes felt by consumers will not be captured in reliability modelling and these limitations will persist even with the new form of the reliability standard. AEMO considers this will require further consideration and ongoing acknowledgement throughout this review.

The VCR approximation does not contradict reforming the standard

With respect to the value consumers place on reliability, the Issues Paper identified the need to have regard for the Value of Customer Reliability (VCR) when developing an appropriate reliability metric, as is required when setting the level of reliability standard. AEMO agrees that is it important to understand the value customers place on avoiding USE and in particular the risk tolerance to tail events.

The VCR approximation[2] does not differentiate between consumer risk and resulting tolerance of different types of USE events, including events of different duration. All USE events are valued equally on a per-hour basis, therefore the VCR assumes societal tolerance for reliability events does not vary based on the duration, scale, or frequency of the event. This is problematic in a future system with significant variability of reliability outcomes, and AEMO agrees with the Issues Paper acknowledgment that the 'VCR may poorly reflect the value of some event types that have not occurred in the past'[3].

[2] VCR is determined by the Australian Energy Regulator (AER) from customer surveys. VCR is subject to review by the AER to be completed by December 2024.

[3] [3] AEMC, Form of Reliability Standard and Administered Price Cap: Issues Paper – page 16: https://www.aemc.gov.au/sites/default/files/2023-03/Review%20of%20the%20form%20of%20the%20reliability%20standard%20and%20APC%20-%20REL0086%20-%20Issues%20paper.pdf

Section 5.2 of the Issues Paper notes the change to the form of the reliability standard should deliver a standard that 'better reflects customer expectations and the value they place on system reliability'. AEMO notes the difficulty of accurately reflecting consumer expectations and a willingness to pay for enhanced reliability. This difficulty is highlighted by the existing contrast between the 0.002% USE reliability standard and the Interim Reliability Measure (IRM) set at 0.0006% USE. The IRM was established on Energy Security Board (ESB) advice, and a subsequent rule change found that maintaining reliability to meet levels of a one-in-10-year summer aligns better with community expectations and provides net positive benefits[4]. AEMO expects that the new form of the standard, developed within the context of the limitations of VCR, should bridge the gap between existing reliability measures and remove the need for multiple standards set at different levels.

It is not AEMO's intention to advocate for change that is not within the scope of this review, rather to highlight that the customer experience of low probability, high impact tail events is likely to be worse than forecast within the limitations of USE and as quantified using the existing VCR. This compounds the need to incorporate modelled tail events in a form of the standard that can be adjusted to reflect customer expectations separately to the VCR approximation.

It is AEMO's view the VCR approximation is that it is just that, and in its proper context the VCR approximation does not contradict improving the standard to account for tail risk. While AEMO notes the upcoming Australian Energy Regulator (AER) review of the VCR is due to be completed by December 2024, AEMO considers there may be value in considering an early review that better aligns timings with this Review of the Form of the Reliability Standard and Administered Price Cap.

For clarity, AEMO supports the Panel's characterisation of reliability risks and seeks to highlight that operational realities, consumer outcomes and the range of risks captured within a forecast tail of the probability distribution are likely to be worse than forecast. This will require further consideration through this review and when setting the level of the reliability standard.

Assessment of reliability metrics

This section provides feedback on the assessment priorities for reliability metrics. In addition to incorporating tail risk and given the difficulties explaining and understanding the existing reliability metric, the review should prioritise developing a new form of the standard that is widely understood and describes outcomes for consumers.

The standard must be tangible to consumers and policymakers

AEMO is generally supportive of the objectives and criteria set out by the Panel (Section 5.2) and acknowledges the difficulty in balancing the key priorities for an amended reliability standard.

AEMO agrees with the Panel's statement 'the current reliability standard is complex to understand and translate into a tangible measure to which consumers can relate'[5]. AEMO has found this to be an ongoing

[4] ESB, Interim Reliability Measures – RRO Trigger: https://web.archive.org.au/awa/20210603165453mp_/https://energyministers.gov.au/sites/prod.energycouncil/files/publications/documents/ESB%20Consultation%20Paper%20on%20Draft%20Rules%20Interim%20Reliabity%20Measures%20-%20RRO%20Trigger%209%20September%202020.pdf

[5] AEMC, Form of Reliability Standard and Administered Price Cap: Issues Paper – page 23: https://www.aemc.gov.au/sites/default/files/2023-03/Review%20of%20the%20form%20of%20the%20reliability%20standard%20and%20APC%20-%20REL0086%20-%20Issues%20paper.pdf

challenge when explaining and working directly with stakeholder groups including government and policy-makers. As a measure of energy, USE does not directly describe customer outcomes and the existing average annual USE is easily misinterpreted as it does not represent an actual outcome, nor reflect what is reasonably expected on a yearly basis.

The continued use of a complex and abstract metric increases the risk of government or policy-makers developing additional reliability policies or standards that they feel better speak to consumer outcomes. For example, the New South Wales Government's 'Energy Security Target' (EST) is defined to set the level of New South Wales Government's reliability expectations[6]. The EST sets the target capacity required to meet forecast New South Wales maximum consumer demand in summer (measured using a 10% probability of exceedance (POE)), with a reserve to account for unexpected loss of the two largest generating units in New South Wales[7].

It is important this review prioritises the development of a new form of the standard that is widely understood and describes outcomes for customers. This would limit the need for additional, and potentially conflicting, jurisdictional targets set above or at a level different to the reliability standard.

Within the Panel's identified criteria of 'clear, simple and implementable', AEMO believes that the modelling and technical feasibility of the reliability standard and the process to review and update the standard should be specifically considered. This would include the ability to rationalise a quantitative basis for setting the level of the standard, including reflecting a level of customer risk aversion or societal tolerance.

AEMO understands this is to be elaborated within the Directions Paper, but cautions about the use of reliability metrics that include an increasing number of abstract coefficients that do not clearly describe customer outcomes. The development of the form of the reliability standard should not be considered a mathematical problem, but rather the development of a practical energy industry outcome. A reliability metric also needs to be able to report against actual USE outcomes, to consider the performance of the metric.

The Issues Paper refers to the General Assessment Principles developed in the Panel's 2021 Review of the Reliability Standard and Settings Guidelines[8] set as:

1. *Allowing efficient price signals while managing price risk*
2. *Delivering a level of reliability consistent with the value placed on that reliability by customers*
3. *Providing a predictable and flexible regulatory framework.*

As set out, the reliability standard and settings should be set a level that balances new entrant resource adequacy with consumer willingness to pay. Notwithstanding the above discussion on the limitations of VCR, the assessment principles identify the need to balance price signals with the level of reliability sought by customers. By design, incorporating tail risk in the reliability standard means that lower probability, higher impact events will seek to be accounted for through higher market settings[9] and the resulting supply resource

[6] NSW Government, NSW Electricity Strategy 2019:https://www.energy.nsw.gov.au/sites/default/files/2022-08/2019_11_NSW_Electricity StrategyDetailed.pdf

[7] AEMO, Energy Security Target Monitor Report, October 2022: https://www.energy.nsw.gov.au/sites/default/files/2022-12/28October 2022-Energy-Security-Target-Monitor-Report.pdf

[8] AEMC, 2021 Reliability Standard and Settings Guidelines: https://www.aemc.gov.au/sites/default/files/2021-06/Review%20of%20 the%20reliability%20standard%20and%20settings%20guidelines%20Final%20Guidelines%20for%20publication%20%28pdf%29.PDF

[9] Or other mechanisms and actions that are guided by the reliability standard.

mix. However, the volatility and difficulty to predict tail events, makes it difficult for new entrant resources to gain revenue certainty when making investment decisions based on a need to enter the market to supply tail events.

The structure of a reliability metric should also be considered for how it will operationalise through the settings and the impact of higher price settings to incentivise the appropriate resource mix and characterise system reliability risk. The Panel should be cognisant of the actions that fall out of a reliability metric, including assessment of the need and as the trigger of out-of-market mechanisms such as the interim reliability reserve (IRR), Reliability and Emergency Reserve Trader (RERT), Retailer Reliability Obligation (RRO). The new form of the reliability standard should be considered in its effectiveness to drive the procurement of resources that the power system requires in the future, and within the context of the changing needs of the power system.

Modelling approach

AEMO is generally supportive of the Panel's proposed modelling approach and agrees with the need to undertake further modelling through this review to continue to analyse the future USE risk profile and develop an analytical basis for comparing potential reliability metrics.

The Issues Paper describes the Panel's proposed approach to build on existing AEMO *Integrated System Plan* (ISP) modelling and overlay new worst-case VRE data, that has been adjusted by matching existing USE outcomes to include expected events such as long-term VRE droughts.

At a high level, AEMO is comfortable with the approach as described in the Issues Paper, including the use of AEMO's ISP modelling as a foundation for future scenarios. The ISP inputs, assumptions and scenarios have been developed and tested, including through industry consultation, and should be considered as the foundation for analysing future drivers of reliability events over time.

As documented in the System Operability Appendix to the 2022 ISP[10], low levels of USE are forecast in the ISP due to the additional capacity built into the plan. Further analysis depicting the impact of 'dark doldrums' weather demonstrates that the average and spread of VRE used is representative of the past 40 years and that the plan is likely resilient to such events. In undertaking analysis on the reliability outcomes that may emerge under a future capacity outlook, the analysis may need to remove some generation from the plan, to identify the form of the reliability risk at a marginal reliability outcome.

AEMO further recognises this as an opportunity for the Panel to use ISP scenarios to develop longer-term and varied outlook scenarios, that are outside the ESOO 10-year horizon and the definitional limitations of what is included in ESOO modelling. As the new form of the standard will seek to commence in July 2028, further analysis of the range of longer-term scenarios should be used to inform its development.

While AEMO agrees with the importance of considering the USE shape to understand and identify the drivers of reliability risk, AEMO has found it increasingly useful to characterise and analyse the dimensions of reliability events as frequency, depth, and duration. Further discussion on AEMO's characterisation of tail risk is included in the following discussion of reliability metrics.

While broadly supportive of the approach, AEMO welcomes the opportunity to work with the Panel throughout this review to ensure the form of the reliability standard is feasible from a modelling and technical perspective.

[10] Section A4.2.4: https://aemo.com.au/-/media/files/major-publications/isp/2022/2022-documents/a4-system-operability.pdf

Reliability metrics

This section discusses potential metrics for the form of the reliability standard, and section also proposes the Panel consider AEMO's preferred approach, a multi metric standard that accounts for tail risk, describes tangible outcomes for consumers, and can be calibrated to a level of average USE (from which the Panel can then estimate the Reliability Settings).

First, it is important to note the existing standard, mean USE, does include tail risk – unlike some others like LOLE or LOLP – because the large tail events do influence the mean USE value. Therefore, AEMO has previously discussed and proposed an average USE value of 0.0006% as used in the IRM as a reasonable proxy for managing tail risk.

A more tangible standard, using depth, duration, and frequency to set mean USE

The Issues Paper discusses the use of multi or hybrid metrics that combine two or more elements that are required to meet the reliability standard, and notes 'individual risk metrics may be able to be combined into a multi metric standard to address a range of different types of risk'.

AEMO agrees with the Issues Paper characterisation and views multi or hybrid options as a reasonable option for the reliability standard, as they:

1) cover the depth, duration, and frequency of reliability events, and in doing so
2) represent tangible outcomes for customers that are easier for customers and policy-makers to understand.

Combining individual metrics into a multi metric standard may best be used to characterise the different elements of USE. AEMO is of the view that considering USE as a function of depth versus duration, instead of energy, is a practical way to characterise reliability risk and relate directly to customer or societal tolerance of tail risk events.

The existing reliability standard defines USE as the percentage of energy demanded that results in loss of load'. AEMO, however, is increasingly analysing and viewing USE as the dimensions that comprise USE outcomes.

Figure 4 below plots forecast average USE outcomes for Victoria in FY24 based on the standard 2,400 simulations used for Victoria in FY24 in the 2022 ESOO but presents them as a function of depth and duration. Each USE outcome (red dot) is plotted based its specific dimensions of duration (x-axis) and depth (y-axis), which is shown as a percentage of average regional demand averaged all the simulations. The forecast annual average USE for this period was 0.0003% USE.

The reliability events expressed as duration versus depth in Figure 4 are translated from USE, and can be related back to the existing average USE reliability standard and IRM shown on the graph as the lines converge towards a smaller depth of event as the relative duration increases. Each dot sitting above the reliability standard and IRM indicates that the event has exceeded the standard.

AEMO considers this graph is a useful way to view USE outcomes and tail risk, demonstrating that deep USE events even at shorter duration, and vice versa, will exceed the reliability standard.

Figure 4 USE outcomes VIC FY24 - average depth versus duration

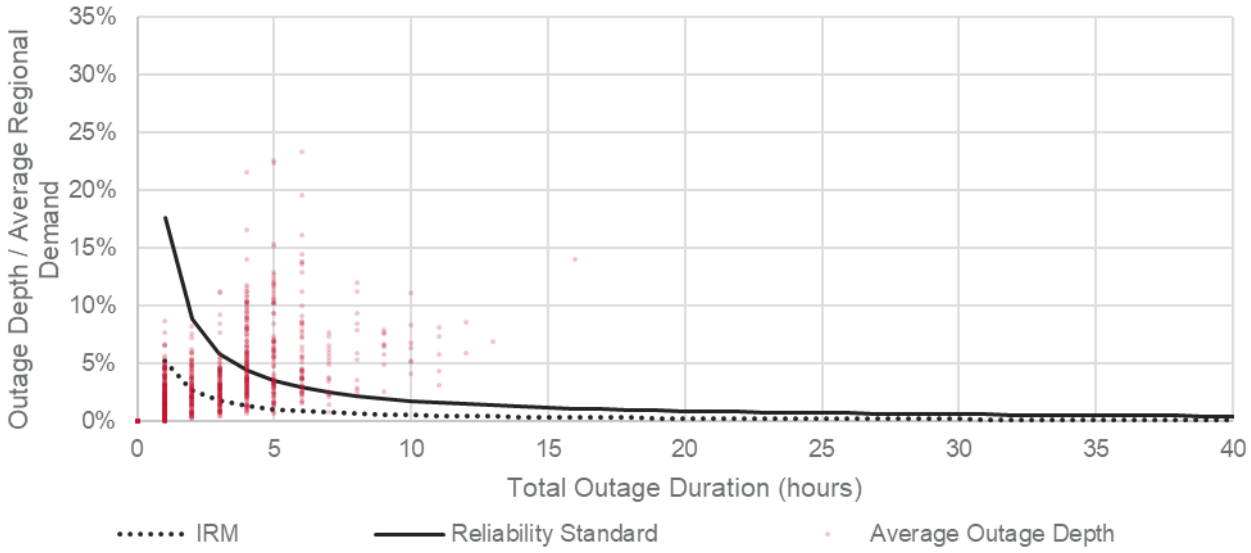
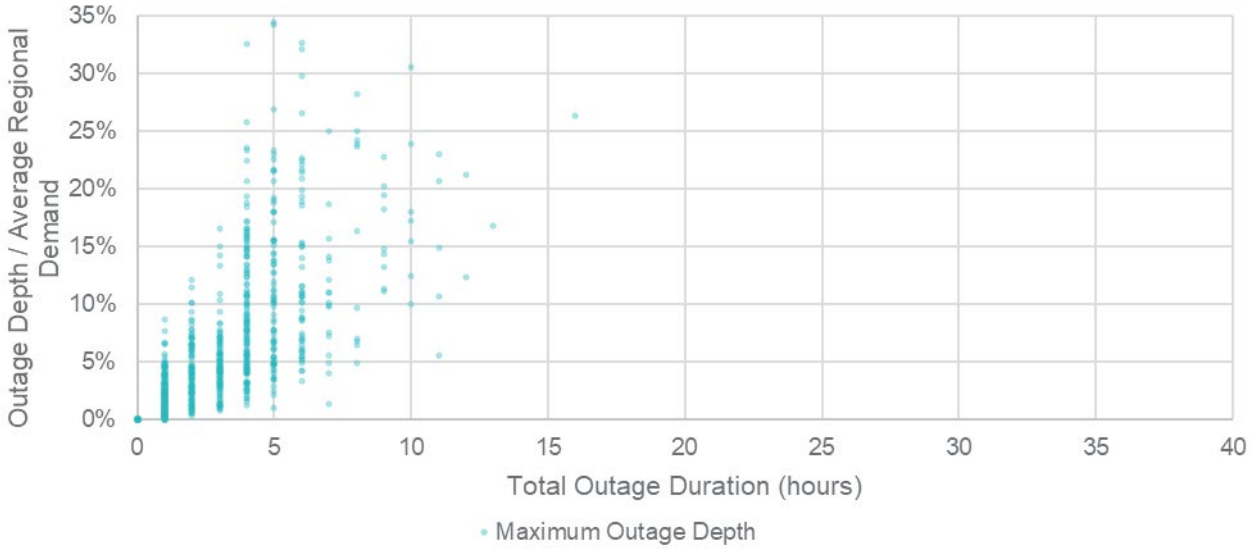


Figure 5 below also plots the maximum outage depth for each outage duration demonstrating the modelled worst-case reliability outcome for each duration. Each USE outcome (green dot) represents the maximum outage depth for each duration across the modelled simulations and, when compared to the average depth (shown in Figure 4 above), highlights how the extent of the tail may be expressed as the dimensions that comprise USE.

Figure 5 USE outcomes VIC FY24 - maximum depth versus duration



The table below also demonstrates how expressing the equivalent depth and duration based on a level of USE translates USE into a specific description of customer outcomes. For example, in Victoria in FY24, there is an 11% probability of an annual average loss of load of >0.0006% USE, equivalent to approximately a one-in-9-year event. This amount of USE can be considered at different depths and durations, as 5.3% of

average regional load shed at 1 hour or 1.3% of average regional load shed for 4 hours (this is estimated as equal to approximately 26,600 households).

The right two columns in the table follow to show average depth and duration of the reliability events above different levels of USE or existing standards. For example, USE outcomes or events above the current 0.0006% IRM have an average duration of 4.56 hours at an average depth of 5.3% of regional load.

Table 1 Average annual USE expressed as depth and duration

	VIC FY24						
	Probability		% of Regional Load Shed		# Households Shed	Outcomes above the standards	
			1 hr	4 hrs	4 hrs	Avg Duration	Avg Depth
% > 0 USE	20%	1 in 5				3.18	3.7%
% > 0.0006% USE	11%	1 in 9	5.3%	1.3%	26,635	4.56	5.3%
% > 0.002% USE	5%	1 in 19	17.5%	4.4%	88,783	5.72	7.5%
% > 0.005% USE	2%	1 in 54	43.8%	11.0%	221,958	7.07	10.2%
% > 0.01 % USE	0%	1 in 439	87.6%	21.9%	443,915	8.22	16.0%

AEMO is of the view that expressing annual average USE as depth, duration, and frequency better describes the customer experience, which then can be set in a standard to reflect societal tolerance to tail events. AEMO considers this a priority and is increasingly characterising USE in this way as part of its ESOO and other forecasting functions.

Multi metric standard calibrated to an equivalent USE at one-in-10-year probability

Based on this characterisation, AEMO proposes a multi metric reliability standard that is expressed as the depth, duration and frequency with the form of the standard as: the average annual outcomes at or above a one-in-10-year probability. For example, this standard could be expressed as:

> *the average annual outcomes that are at or above a one-in-10-year probability may not be greater than x% of average regional load shed for 4 hours, or equivalent.*
>
> where:
>
> Measure = mean annual average USE, like used in current standard
>
> Frequency = 1 in 10
>
> Depth = x% of average regional load
>
> Duration = y cumulative annual hours

This form of the standard captures tail risk by describing annual average reliability outcomes in terms of the level of individual high impact, low probability tail events that would seek to be limited by maintaining reliability to the level of the standard. This metric defines within the standard the **depth** (percentage of average regional load) and **duration** (cumulative annual hours) and limited at the end of the probability distribution at a one-in-10-year probability, **frequency**. By describing the average allowable outcome for consumers within the standard, this allows consumers to manage their societal expectations or allows the standard to be updated and set at a level that meets and directly speaks to the appropriate level of customer risk aversion.

Further, it is useful to note that while expressed in terms of depth, duration and frequency, this approach is statistically equivalent to the Panel's identified CVaR approach, but without the weight parameter included in the composite straw-person metric. By expressing allowable reliability levels at a one-in-10-year probability, this metric is describing average annual outcomes of the tail events at a 10% CVaR. If the Panel wanted to retain the existing annual average USE metric, AEMO considers that this more tangible approach to tail risk could be used to replace or be used in addition to the existing metric.

AEMO considers that this approach bridges the gap between the complexity of the concepts of USE and removes the need for a weighting parameter as included in the Panel's straw-person proposal, by focusing on describing the consumer outcomes to the level that effectively reflect societal tolerance to reliability risk. This characterisation and approach is likely to be tangible to policy-makers as they consider how USE would be felt by customers.

Also, as this characterisation is equivalent and directly translated from the existing measure of USE, AEMO views this as a practical and implementable approach, with USE still to be used to model, forecast and measure outcomes using established practices and forecasting guidelines. This is therefore not a significant divergence from existing processes, but rather a translation that may be utilised to describe the customer experiences of reliability outcomes and aligns with the need to reflect the level of customer risk aversion to tail risk within the reliability standard.

The use of average outcomes at or above a one-in-10-year probability within this metric also acknowledges the impossibility of guaranteeing a maximum allowable level of a tail event. In this way it retains a useful characteristic of the current form of the standard. The Issues Paper includes consideration of estimating the maximum USE as a potential metric for the reliability standard (Table 6.1). While setting a level of maximum USE has the advantage of being easy to understand and directly reflect the level of customer risk aversion, AEMO notes that it is impossible to guarantee a maximum allowable level of reliability, particularly within the context of the discussed USE and VCR limitations. The use of average USE outcomes within this proposed metric is a better reflection of the operational reality of maintaining reliability to meet forecast outcomes.

While the Panel's review is at a relatively early stage, with further analysis to be undertaken, AEMO considers this metric as its preferred approach at this stage of the review process and looks forward to the opportunity to discuss this with the Panel further. AEMO considers the Panel could juxtapose this simpler, more tangible standard, framed around depth, duration and frequency, against the more complex straw-person identified in the Issues Paper.

Alternative options

Mean USE (the existing form of the standard)

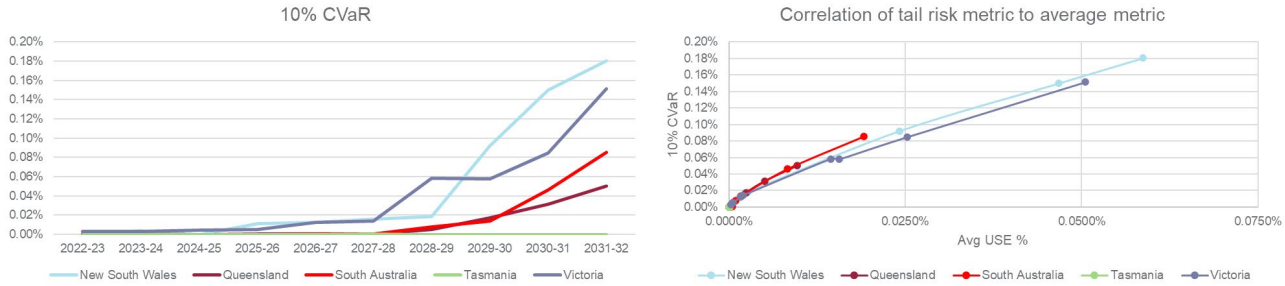
The existing form of the reliability standard is included for consideration. While AEMO agrees with the Issues Paper characterisation that the existing average annual USE reliability standard has the potential to mask the impact of tail events, it is important to acknowledge that average annual USE is correlated to tail risk because tail events are included in the calculation of the average USE amount. It is the averaging out of the USE value caused by the number of simulations with no forecast USE that masks the impact tail risk, not the form of the standard – the average annual USE value can be calibrated to account for tail risk.

The problem with "averaging out" is that consumers will not experience an average value, and to aim for an average of 0.002% USE is to accept that some large reliability events will occur. Therefore, while AEMO

considers it important to acknowledge mean USE is inclusive of tail events, expressing the standard as the depth, duration and frequency provides a clearer limit to consumer outcomes.

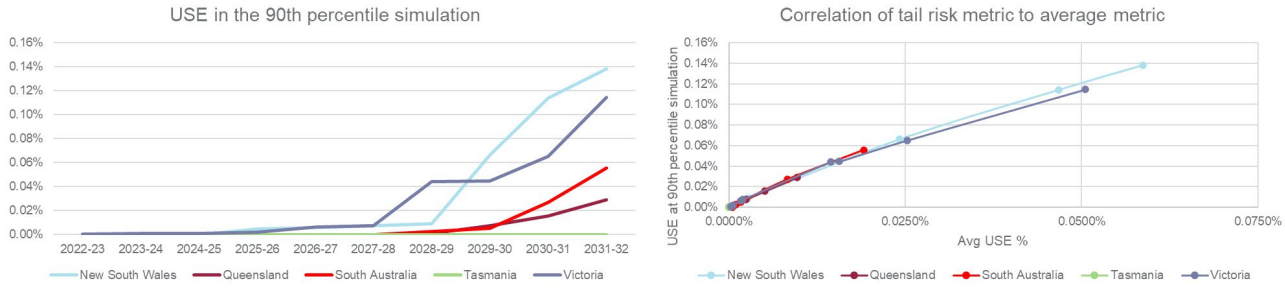
Figure 6 shows how USE is correlated to tail risk for the modelled results in the Update to the 2022 *Electricity Statement of Opportunities*. The figure presents the CVaR for the top 10% of simulations with USE (or the average USE of the top 10% of simulations, shown as the percentage of USE), compared with the average USE metric. The USE resulting from the top 10% of simulations dominates the mean, hence there is a high correlation (shown on the right), although of course it is not 1:1 because there are other simulations with USE. This interrelationship clearly shows tail risk is included for in the average annual USE value, and suggests this could be retained but calibrated lower to account for tail risk.

Figure 6 Correlation between CVaR (10%) and average USE (Update to the 2022 Electricity Statement of Opportunities)



As a comparison, Figure 7 below picks the USE results from a single simulation that represents the 90th percentile USE. By selecting the 90% value, which is quite high, and then comparing with annual average USE, this figure is of course less correlated than Figure 6 because it does not include as much data. The correlation (below right) is still relatively consistent, demonstrating what we know, that generally higher events will be relative to a higher average annual USE.

Figure 7 Correlation between USE in 90th percentile simulation and average USE (Update to the 2022 Electricity Statement of Opportunities)



As such, while average annual USE is not considered to specifically describe tail risk events, tail risk and higher events are correlated and included within the average annual USE measure. While average USE is used in these graphs to demonstrate the correlation between average annual USE and tail risk parameters, this metric or analysis does not overcome the USE or CVaR communication challenge.

The use of average USE as the basis of the multi metric standard based on frequency, duration and depth allows established modelling and analysis such as this to form the basis of the standard, that is then expressed in a clearer and more tangible way for consumers.

Conditional Value at Risk

The composite metric identified by the Panel including expected USE and conditional USE (or CVaR) is an example of a multi metric standard. While AEMO agrees that it effectively incorporates tail risk into the standard, it adds a further layer of complexity to the standard, and is therefore unlikely to satisfy the requirement for the standard to be simple and represent tangible outcomes for consumers. This is particularly so when stakeholders already struggle with understanding the expected reliability from the average USE.

$$R_{USE} = w \cdot \text{expectedUSE} + (1- w) \cdot 95\%\text{-CVaR}_{USE}$$

The composite standard combines the existing expected USE value with CVaR as the mean of the *a*% (in this case 5%) worst cases (the "tail") of the USE probability density function. The CvaR tail risk indicator describes the extent of the tail, while the *w* weighing parameter seeks to reflect the level of risk aversion and described as suitable for adjusting in response to evolving conditions.

While the weighting parameter *w* intends to be adjustable to reflect a level of customer risk in response to evolving market conditions, it is unclear what would be an appropriate basis for determining this value. Extensive modelling would be required to demonstrate how *w* and *a* combine to efficiently reflect societal tolerance and there is a risk that this theoretical composite standard would be based on overly complex modelled simulations, not real-world outcomes.

Further, AEMO considers it would be significantly more difficult to explain to governments and policy-makers, and it could be difficult to justify the use of CVaR as a reliability standard. For example, it is not clear what the customer outcome is or how it relates to the customer experience of USE.

The abstract nature of CVaR also limits its ability to report actual USE outcomes and measure against the performance of the reliability standard ex-post. In practice there are no defined "worst case" outcomes as per the forecast resulting in a comparison of actual values to the composite metric. AEMO requires the ability to review and consider the performance of the standard in Medium Term Projected Assessment of System Adequacy (MT PASA) and ESOO cycles, and the inability to review and consider if the reliability standard was met is likely to be a significant inhibitor in effectively setting the standard going forward.

Further, the close correlation between the mean USE value and the tail value of USE events indicates, instead of implementing a more complex methodology like CVaR, it would be easier to simply reduce the mean USE value to account for tail risk, for example using 0.0006%, rather than 0.002%. This leads AEMO to believe the Panel should investigate simpler, more tangible metrics that better describe the nature of reliability events and their likelihood, rather than more complex variations to the existing standard.

As such, AEMO considers that the composite standard of average USE and CVaR is not fit-for-purpose, relative to the multi metric standard that balances the efficiencies of average USE, with the need to prioritise simplicity and consumer outcomes. While the composite standard is beneficial in describing both the total expected USE and the extent of the tail, the combination and use of additional parameters results in a significant increase in complexity that is unlikely to be acceptable to stakeholders. If the Panel prefers to retain the existing average annual mean USE, the multi metric standard can be used in addition to the existing metric, without the need for a joint composite metric.

Largest annual outcome tolerable at a one-in-10-year probability

A variation to the approach above for consideration is to take a form of the reliability standard that seeks to maintain reliability to the largest annual outcome tolerable at a one-in-10-year probability. For example, this standard could be set and described as:

> *The annual outcome at a one-in-10-year probability may not be greater than 0.002% USE.*

The aim of this is to focus not on the average outcome, but on the maximum tolerable as a certain probability. While AEMO acknowledges this variation is appealing to limit allowable reliability outcomes to a maximum level, it is problematic to use and set expectations of a 'maximum tolerable' level from within a reliability standard. As discussed above, the reliability standard is unable to guarantee a maximum loss as operational circumstances reliability outcomes do not eventuate as forecast.

Although included for consideration, AEMO would consider this metric or a standard that seeks to describe a "maximum" value to be unsuitable, because it misrepresents the standard as being able to guarantee a maximum loss, which is untrue.

The standard could also include predetermined conditions

A further variation to the multi metric standard described above is to include additional predetermined conditions to the standard that limit average allowable USE outcomes under defined circumstances. This multi metric approach is akin to combining a probabilistic measure with a deterministic one to form the reliability standard. For example, the standard could be set and described as:

> *Expected USE may not be greater than 4.4% of average regional load shed for 4 hours, or equivalent, under 10% POE demand conditions and 90% POE VRE conditions.*

This is beneficial as it allows for consideration of scenarios that are not picked up in simulation modelling (probabilistic) with the use of a deterministic measure that may be set relative to the level of customer risk aversion or societal tolerance to tail events.

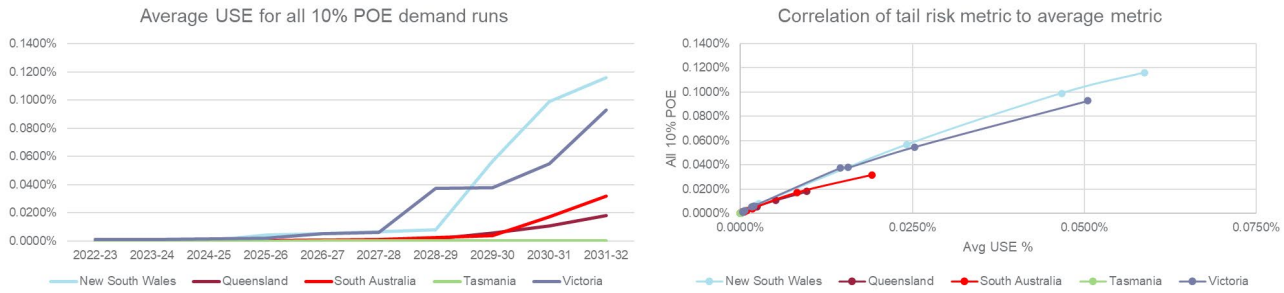
The usefulness of this standard depends on whether the predetermined conditions are the predominant cause of reliability events. Reliability events may be caused by multiple different types of conditions, and a deterministic standard may need to include a 90% POE VRE forecast, the loss of the largest two units, and a 10% POE demand forecast. The need to include additional conditions to cover the range of scenarios and drivers of reliability events may lead to an increasingly high level of reliability that is inherently difficult to maintain under the defined conditions. Reliability risk is increasingly characterised across of a variety of VRE, energy-limited resource and demand outcomes, and meeting a reliability standard that requires worst case across multiple different conditions, while accounting for tail risk, would likely result in an unreasonably high and inefficient level of reliability.

Returning to a similar comparison as shown above in Figure 6 and Figure 7, it is clear that if the Panel wants to use a metric like average USE, but calculate it including a predetermined condition, like high demand, then the correlation is broken somewhat when compared to the strong correlation to CVaR. This is shown in Figure 8 below (right) where the conditional tail risk metric of average USE at 10% POE demand is shown relative to average USE. While the correlation is not as strong as Figure 6 (CVaR versus average USE), it remains strong overall. This demonstrates that:

- 10% POE demand is a significant contributor to USE outcomes and that most, but not all, USE outcomes occur during 10% POE demand periods, and

- 10% POE demand is a relevant predetermined condition, and may be superior to another predetermined condition, such as the loss of the largest two units, although this would need to be tested.

Figure 8 Correlation between average USE at 10 POE demand and average USE – Update to the 2022 ESOO



AEMO recommends this metric for conceptual consideration, but is unsure if it is fit-for-purpose given the limitation of predetermined conditions as the drivers for changing reliability events.

Administered Price Cap

AEMO notes the Panel is also considering changing the APC. The imposition of a price cap reduces the incentives for supply and therefore, irrespective of the setting of the APC, supply cannot be guaranteed. Ideally the APC needs to be sufficiently high to encourage participants to make plant available for dispatch during administered pricing periods, and the higher the APC is, the more likely it is to do so. However, should the scarcity conditions persist after breaching the cumulative price cap and entering administered pricing, there remains the chance the APC is insufficient to encourage efficient dispatch. This is the reality of imposing a price cap.

The APC could be adjusted to try to account for a range of circumstances, such as high fuel costs and increases in CPI, but it is unlikely, particularly in a system that may be affected by energy shortages and opportunity costs, that the circumstances can be reasonably foreseen. So, while the Panel should consider how the APC could be adjusted to better account for prevailing operating costs, it should also recognise that amending the APC is not the only option to encourage dispatch. In particular, the Panel should be cognisant that Participants are entitled to claim compensation from the AEMC under administered pricing.

Being a cap, ideally the APC should be above the prevailing operating costs of most plant. This puts pressure on the Panel to increase it further, but the problem with increasing the APC substantially is that it reduces the effectiveness of it in mitigating participant exposure to high price events.

AEMO looks forward to the Panel's analysis of the usefulness and ease of implementation of each of the proposals suggested in the Issues Paper[11]. It is sensible that the Panel is considering the opportunity to set the APC higher under certain conditions, however the Panel should investigate whether this could encourage bidding behaviour that aims to force an increase to the APC.

The APC should be set at a level sufficiently high to incentivise generation to participate during emergency situations. AEMO considers that the transitional increase from $300/megawatt hour (MWh) to $600/MWh that is applied until 30 June 2025, and the subsequent APC to 2028 of $500/MWh, should be sufficient at least in the short term to allow for more orderly market operation during those periods and relative to June 2022.

AEMO does not consider the APC affects long-term commercial decision making – the Reliability Standard, Market Price Cap and Cumulative Price Threshold have this effect, particularly if the Standard and Settings are amended to encourage the market to invest to meet the changing distribution of reliability events. Similarly, AEMO considers changing the APC, by say a couple of hundred dollars, would have little to no effect on the ability for parties to enter forward contracts and power purchase agreements.

[11] Reliability Panel, Review of the Form of the Reliability Standard and Administered Price Cap Issues Paper – Page 38 - https://www.aemc.gov.au/sites/default/files/2023-03/Review%20of%20the%20form%20of%20the%20reliability%20standard%20and%20APC%20-%20REL0086%20-%20Issues%20paper.pdf