make history.

THE UNIVERSITY of ADELAIDE

Review of FORECAST ACCURACY METRICS

AEMO Forecasting Reference Group Presentation

Review of FORECAST ACCURACY METRICS

- AEMO annually produces forecasts for each region of operational energy consumption, extreme half-hourly demands, and available supply
- Forecast accuracy is reported annually in the annual Forecast Accuracy Reports (Annual Reports)
- Annual Reports build confidence in AEMO forecasts and *inform the continuous improvement of the forecasts*
- This review:
  - follows on from the 2019 Review of Forecast Accuracy Metrics, also by The University of Adelaide
  - assesses the forecast accuracy metrics that are described in the “Forecast Accuracy Report methodology” paper dated August 2020 and presented in the Annual Reports
  - considers in detail the 2020, 2021 and 2022 Annual Reports
  - notes that many of the recommendations of the 2019 Review have been implemented and have significantly improved the communication of the forecast accuracy

THE UNIVERSITY of ADELAIDE

Review of FORECAST ACCURACY METRICS

- We found the “Forecast Accuracy Report methodology”, dated August 2020, to be an excellent description of the relevant issues and approaches
- The quality of the accuracy reporting methodology and the Annual Reports is very high and we commend AEMO on their standards and progress
- We provide 26 recommendations from our review (6 + 5 + 13 + 2)
- These recommendations are based on the high-level information available to the review team
- AEMO will need to assess the feasibility and benefits of implementing each recommendation
- Specific details around what information should be presented or prioritised is to be determined internally by AEMO and informed by stakeholders

THE UNIVERSITY of ADELAIDE

General Recommendations

General Recommendations

Recommendation 1: Continue with the use of the forecast categories and reporting methodologies described in Table 4 of the “Forecast Accuracy Report methodology”, dated August 2020

	Deterministic	Probabilistic
Input forecasts	**Simple percentage error metrics are most appropriate:** e.g. growth, new connections, losses, …	**Qualitative description of accuracy may be most appropriate:** e.g. weather impact on behaviour, transmission, generator output, …
Component-based output forecasts	**Best to assess the contribution of each input to aggregate accuracy:** e.g. operational energy consumption	**Challenging to assess accuracy using a single observation, requiring exploratory analysis and qualitative justification:** e.g. extreme demand, traces, DSP, supply availability, …

THE UNIVERSITY of ADELAIDE

General Recommendations

Recommendation 2: Continue with the use of forecast accuracy reporting as a tool to drive improvements in the forecasting methodology

Recommendation 3: Continue with the use of Definition 3 for percentage error

$$\boldsymbol{percentage\ error} = \frac{\boldsymbol{forecast - actual}}{\boldsymbol{actual}} \times 100$$

a positive error means that the forecast was higher than the actual

It is the more easily interpreted definition given the framing of the report as assessing the accuracy of the forecast against the actual

Ensure that this framing is used consistently throughout

THE UNIVERSITY of ADELAIDE

General Recommendations

Recommendation 4: Consider providing information for each row in Table 1 of each Annual Report indicating the *assessability* of that metric, using the three categories (strongly, moderately and weakly) as defined in Section 2.2.1 of each Annual Report

Table 1 Forecast accuracy summary by region, 2021-22

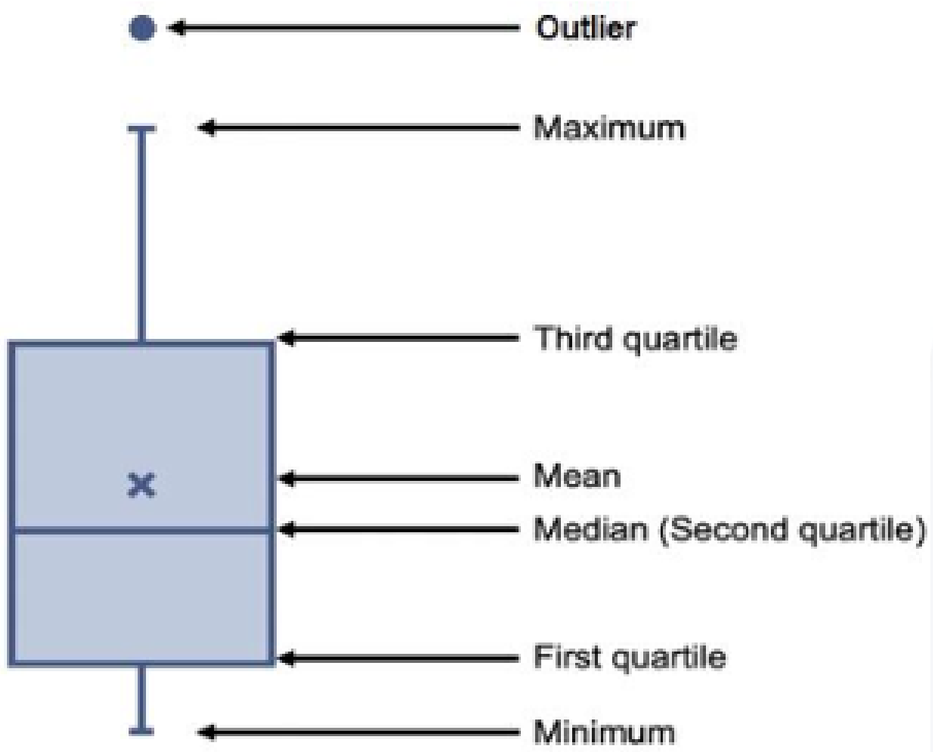
Forecast Component	NSW	QLD	SA	TAS	VIC	Comments
Energy consumption	🟡	🟡	🟢	🟢	🟡	Two regions were within 3%, but New South Wales, Queensland and Victoria all had larger variances, mainly driven by an issue with forecasting underlying business mass market consumption. A fix was provided in the *Update to the 2021 ESOO* in April 2022.
Summer maximum demand	🟢	🟡	🟡	🟢	🟢	Good alignment across most regions except Queensland and South Australia. Actual maximum demand in Queensland fell just above the forecast range, driven by very high humidity well into the evening. South Australia had an unusually low maximum demand outcome driven by very mild weather across the summer due to the La Niña climate conditions and fell just under the forecast range.
Winter maximum demand	🟢	🟡	🟢	🟡	🔴	Winter maximum demand outcomes in Queensland and Victoria were above forecast distributions and Tasmania below. While the reasons for the deviation for Queensland and Tasmania are known, a review of how the initial year of the forecast is set is required.
Annual minimum demand	🔴	🔴	🟢	🟢	🟢	Three regions showed good alignment, while New South Wales' actual minimum demand was lower than the forecast distribution and Queensland's actual minimum demand was above the forecast. With distributed photovoltaic (PV) installations showing good alignment, the forecast model requires review.
Demand side participation	🟢	🟢	🟢	🟢	🟢	Generally, less demand side participation (DSP) was observed than forecast but with too few observations to conclude with certainty. Queensland had many high price periods though and on average more observed DSP than forecast, but during peak demand, it showed good alignment.
Installed generation capacity	🟢	🟡	🟢	🟢	🟡	New generator installations matched expectations in Tasmania and were relatively close in New South Wales and South Australia. Start date delays and slower commissioning in Victoria and Queensland significantly impacted availability compared with what was modelled.
Summer supply availability	🟡	🟡	🟢	🟡	🟢	Extended outages of key units during summer months together with lower summer availability resulted in actual availability being lower than what was modelled for New South Wales, Queensland and Tasmania, noting some units may not have been made available due to low levels of supply scarcity.

THE UNIVERSITY of ADELAIDE

General Recommendations – Box plots

Recommendation 5: Rewrite the description of a box plot to prevent potential confusion between outliers and the maximum/minimum

Figure 3 Explanation of box plots used in this report



THE UNIVERSITY of ADELAIDE

General Recommendations – weather

Enhanced representation of weather in the descriptions and the models

- Weather conditions in the explanation of the extreme demand forecasts are very informative
- Largely based on air temperature at a particular **time** and **location**
- Often elaborated -- more nuanced picture of the weather on that **day** across that **region**

- Specific ideas for enhanced representation of weather include:
  - Use of a demand-weighted spatial average of temperature in each region. This is probably most likely to be relevant in Queensland where notable demand stretches across a very large region with quite different conditions
  - Use of the concept of a heating/cooling load up to that point in time on that day as often the accumulation appears to be more important than the actual temperature at that time
  - Use of the wet-bulb or apparent temperature to include the effects of humidity

Recommendation 6: Consider the introduction of enhanced representations of weather in the descriptions and the models to enable a more rigorous analysis of accuracy

THE UNIVERSITY of ADELAIDE

General Recommendations – benchmark?

Stakeholders would have increased confidence in the AEMO forecasts if their accuracy could be benchmarked against the forecast accuracy of similar organisations. We are not aware of any other jurisdictions with the same requirements in terms of forecasting and reporting on the accuracy of those forecasts.

Recommendation 7: Consider opportunities to benchmark the accuracy of the forecasts against other organisations. In particular, consider benchmarking against the forecasts of consumption from the relevant Distribution Network Service Providers

THE UNIVERSITY of ADELAIDE

General Recommendations – multi-year

- In such a dynamic industry, the forecasting methodology is regularly being updated to seek improvements in accuracy and capture structural changes
- Looking back at the forecasts made 10 years ago to assess their 10-year accuracy may seem to be rather meaningless
- We understand that the 2-year and 4-year timeframes are particularly important
- Do not believe that AEMO should repeat the entire accuracy analysis at 1, 2 and 4 years

Recommendation 8: Consider introducing 2-year and 4-year assessments of the accuracy of certain key elements of the forecasts in the annual reporting process

THE UNIVERSITY of ADELAIDE

Operational Energy Consumption Recommendations

Operational Energy Consumption

Recommendation 9: Continue with the use of percentage error and percentage impact on forecast of total consumption and with the use of tables and waterfall diagrams to represent them

Figure 7 NEM operational as generated energy consumption variance by component

Table 8 NEM operational energy consumption forecast accuracy by component

Category	2021 forecast (gigawatt hours [GWh])	Actual (GWh)	Difference (%)	Indicative impact on total consumption
Cooling degree days	5,177	4,126	25.5%	0.6%
Heating degree days	8,211	8,975	-8.5%	-0.4%
Connections growth	841	559	50.3%	0.2%
Large industrial loads	45,123	43,565	3.6%	0.8%
Rooftop PV	17,520	16,746	4.6%	0.4%
PV non-scheduled generation	2,016	1,742	15.7%	0.1%
Other non-scheduled generation	4,088	4,407	-7.2%	-0.2%
Network losses	9,300	10,257	-9.3%	-0.5%
Operational sent-out	**169,480**	**178,329**	**-5.0%**	**-4.7%**
Auxiliary load	8,562	8,162	4.9%	0.2%
Operational as generated	**178,042**	**186,491**	**-4.5%**	

THE UNIVERSITY of ADELAIDE

Operational Energy Consumption

Recommendation 10: Replace equation 2 on page 17 of the “Forecast Accuracy Reporting Methodology” paper dated August 2020 with the equation:

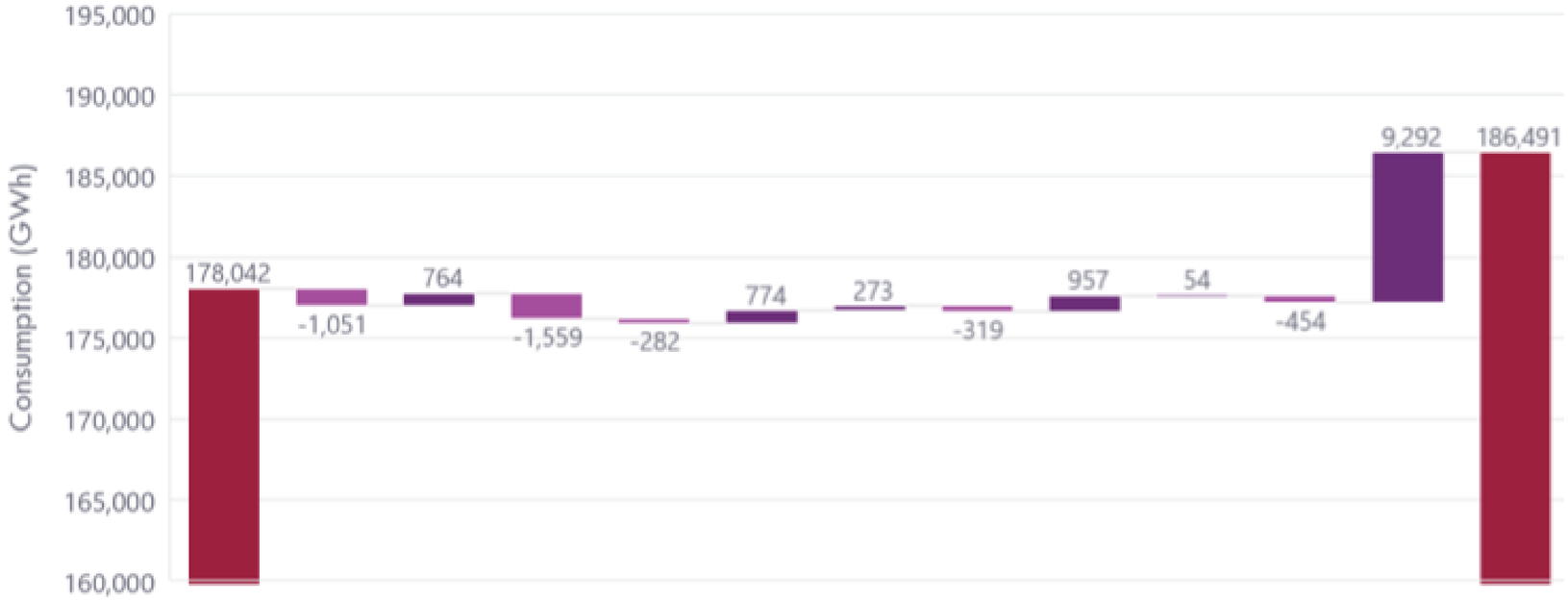
$$error\ from\ forecast\ component = input\ coefficient.\ (input\ \mathbf{forecast} - input\ \mathbf{actual})$$

Recommendation 11: Reorder all the waterfall diagrams to be consistent with the new equation 2 so that each waterfall figure starts with the Actual and presents all the component errors in the forecast that are required to reach the Forecast

THE UNIVERSITY of ADELAIDE

Operational Energy Consumption

Figure 7 NEM operational as generated energy consumption variance by component



- Heating Degree Days and Other Non-Scheduled Generation are both underestimated
- Heating Degree days appears as a positive
- Other Non-Scheduled Generation appears as a negative

Recommendation 12: Ensure that the order and labels of the components in all waterfall diagrams and associated tables are consistent. Consider if it is appropriate to present the three supply-side components (that require a reverse of sign) first in each table and waterfall diagram

THE UNIVERSITY of ADELAIDE

Extreme Demand Recommendations

Extreme Demand

- Forecasting of extreme demand is challenging:
  - tails of a distribution are highly variable
  - unlike more predictable measures, such as the mean
- Forecasting *minimum* demand is significantly harder than forecasting *maximum* demand:
  - lose some of the benefits of the Law of Large Numbers
  - factors that generally might be quite minor can suddenly be quite significant
- Extreme demands are therefore highly variable:
  - discussion-based approach to reporting the accuracy is highly appropriate
  - displayed figures are highly meaningful and informative

THE UNIVERSITY of ADELAIDE

Extreme Demand

Recommendation 13: Continue the use of a discussion-based approach and the use of figures that provide meaningful information about the distribution and drivers of the forecast

Figure 14 New South Wales simulated extreme event probability distributions with actuals

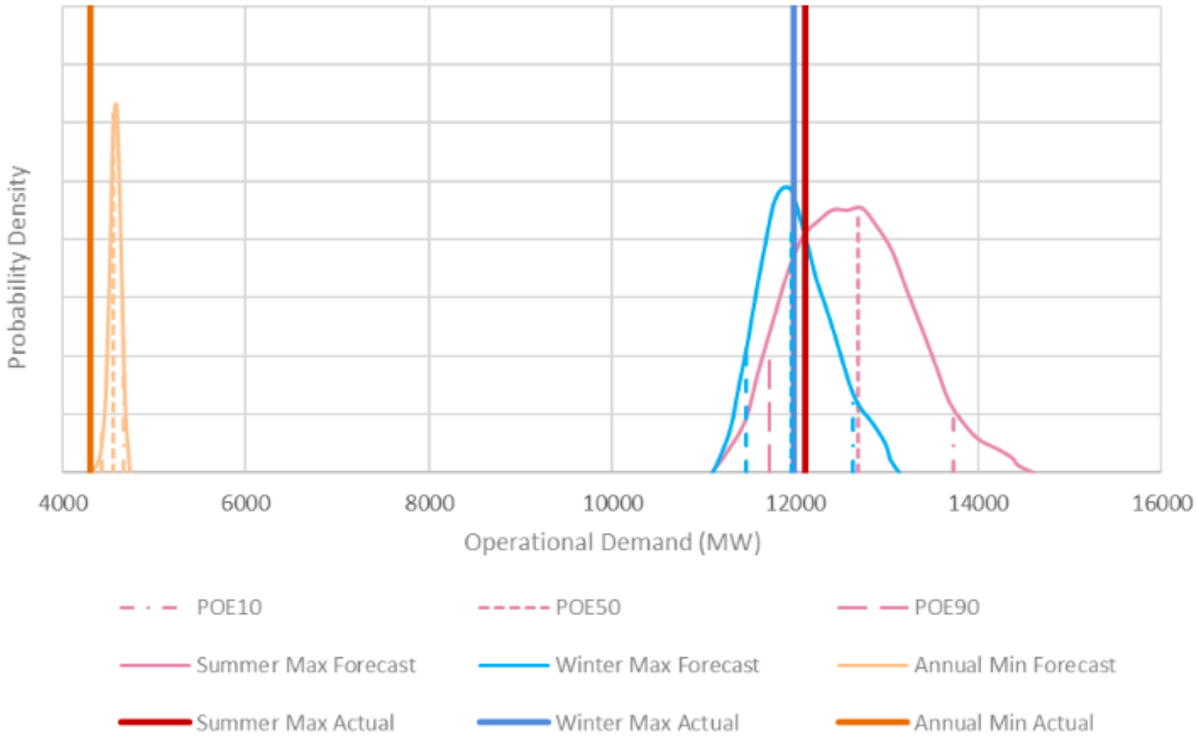
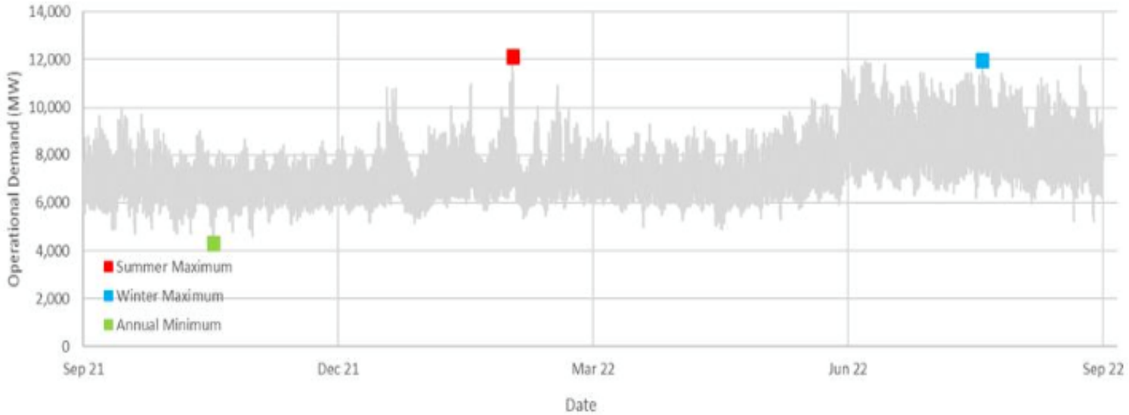


Figure 13 New South Wales demand with extreme events identified

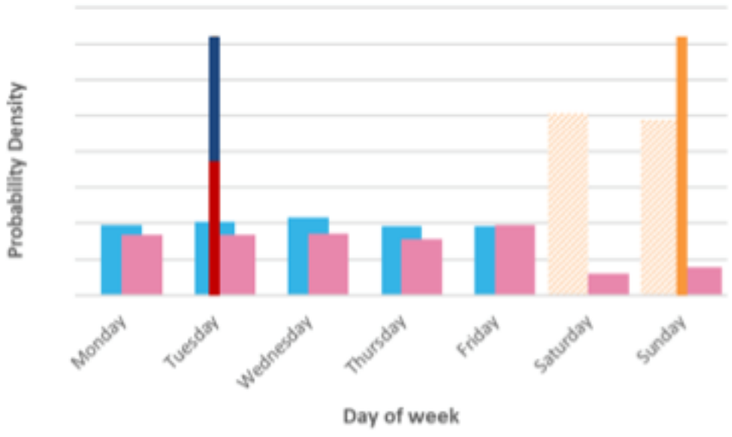


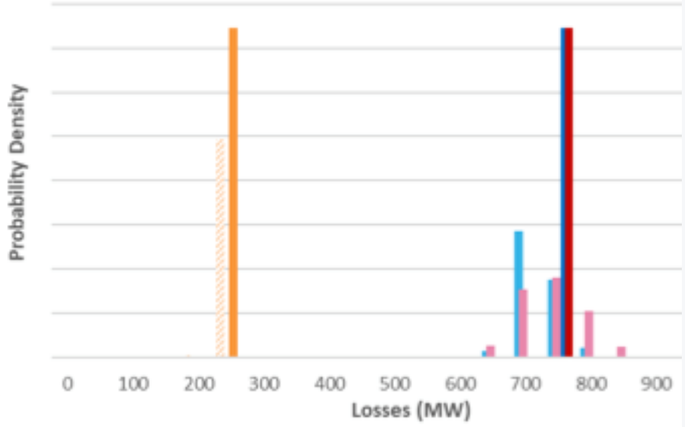
THE UNIVERSITY of ADELAIDE

Extreme Demand – forecast drivers

Figure 15 New South Wales simulated extreme event probability distributions with actuals

Recommendation 14: Retain Figure 15 in future Annual Reports. Review the choice of driving parameters that are displayed and how each parameter is presented

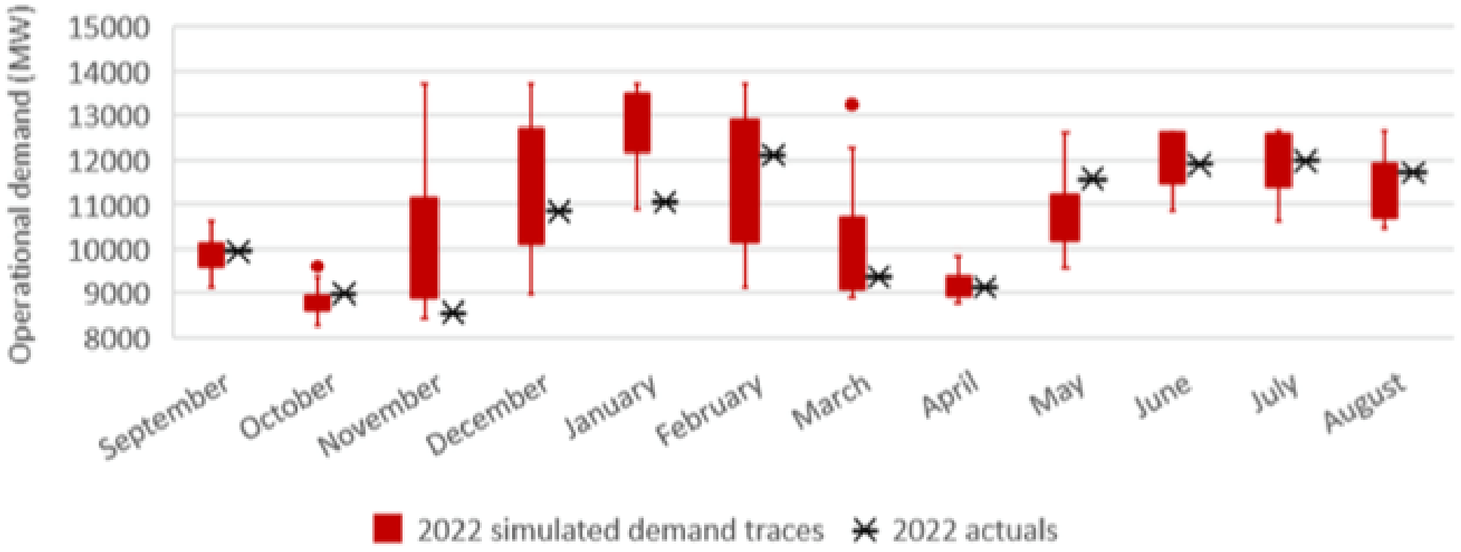




THE UNIVERSITY of ADELAIDE

Extreme Demand – monthly performance

Figure 16 New South Wales monthly maximum demand in demand traces compared with actuals

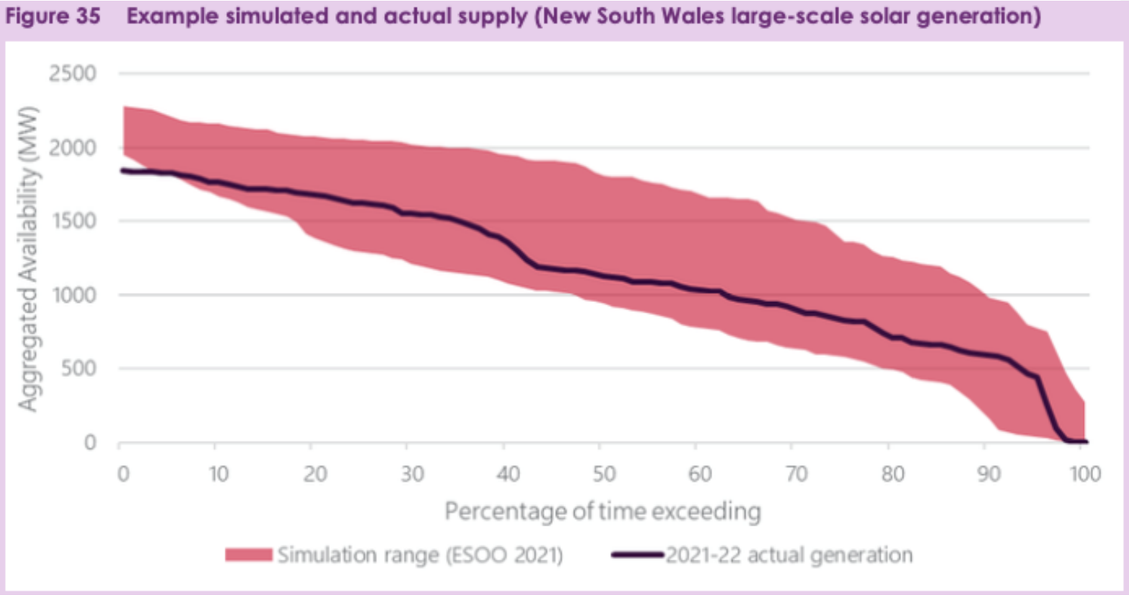


Recommendation 15: Consider whether it would be more appropriate to provide the monthly maximum demand figures based on only 10%, 50% or 90% POE traces, or provide them based on the combination of the 10%, 50% and 90% POE traces all together

THE UNIVERSITY of ADELAIDE

Supply Recommendations

Supply - availability



Forecasting supply availability is crucially important but also highly challenging given:

- the possibility of major outages in individual, highly-significant generators, and
- the uncertainty of the timing of operation of planned/newly-installed generators

The figures used are effective at representing that information on the most important days

Recommendation 16: Continue using the figures for total availability and component generation for each region

THE UNIVERSITY of ADELAIDE

Supply - availability

- If 100% of simulations were shown, the band would become so broad and jagged as to be much harder to interpret
- Restricting the graph to the central 95% is a commonly used and entirely appropriate approach
- Maintaining interpretability is essential, but other methods that are designed to achieve this same end could also be explored, for example:
    - shading each simulation line with an intensity reflecting its centrality among the suite of simulations
        - the more extreme simulations would be very pale and probably quite separated from each other
        - the central simulations would be darker and close together

Recommendation 17: Restricting the graph to the central 95% is a commonly used and entirely appropriate approach. However, consideration could also be given to other approaches that are designed to achieve a similar degree of interpretability

THE UNIVERSITY of ADELAIDE

Supply – generation count and capacity

Table 22 Forecast and actual generation count and capacity, February 2022

New South Wales	Facilities actually operating		Facilities forecast to operate		Difference in capacity (forecast-actual)	
	Count	MW	Count	MW	MW	%
VRE generation	44	4,452	42	4,577	-125	-3%
Non-VRE generation/storage	50	14,731	52	14,781	-50	0%
All generation	94	19,183	94	19,358	-175	-1%

Recommendation 18: Forecast and actual generation count and capacity tables should be restructured so that forecasts are provided to the left of the actuals, and the comparison column is calculated as (forecast – actual) so that the final column follows the generic definition and interpretation of percentage error

THE UNIVERSITY of ADELAIDE

Supply – generation support across regions

- Some regions provide support to other regions during peak demand events, eg
  - Tasmania is a winter-peaking region and the availability of surplus generation provides important support to the mainland during summer peak demand events
  - Queensland provides generation support to New South Wales during high-demand periods in New South Wales
- Insightful to investigate the accuracy of supply forecasts in the supporting region based on the top 10 hottest days in the supported region (or regions)

Recommendation 19: Consider providing an equivalent analysis of the accuracy of supply forecasts in the most important supporting regions based on the top 10 hottest days in the supported region, where appropriate

THE UNIVERSITY of ADELAIDE

Supply – variable renewable energy

- Accuracy of Variable Renewable Energy (VRE) generation forecasts is presented by comparing modelled *availability* to actual *generation*
- All other generation types are presented by comparing modelled *availability* to actual PASA *availability*
- In general, the difference between the two for VRE is assumed to be minimal
- However, on two occasions (Victoria in 2022 on page 80 and Queensland in 2021 on page 65) significant excursions from this are noted

Recommendation 20: Consider whether the accuracy of VRE generation forecasts could be presented consistently on a *generation* basis through both the forecast and reported data

THE UNIVERSITY of ADELAIDE

Supply – demand side participation

- The definition of the trigger events is potentially confusing
- For example, the event set “>$300/MWh” includes all triggers no matter how high the price
- Mix of low, medium and high prices and the median DSP response is most likely corresponding to a **medium** price event
- Bias/variance tradeoff in the estimate of the DSP response:
    - Smaller categories reduce bias
    - Larger categories reduce variance
    - Overlapping categories increases bias
    - Current approach minimises variance but at the expense of significant bias

- Split the trigger events into **disjoint** categories eg “>=$300/MWh AND <$500/MWh”
- Reduce the number of categories

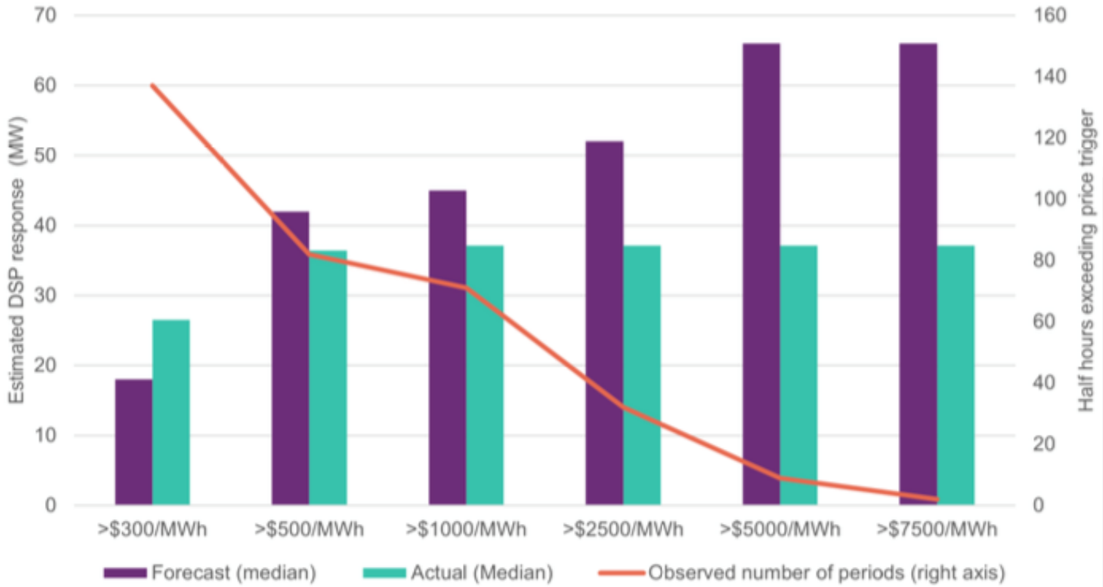
Recommendation 21: Consider modifying the trigger categories into **disjoint** categories and reducing the number of categories to ensure sufficient events in each (disjoint) category while maintaining signal and interpretability

THE UNIVERSITY of ADELAIDE

Supply – demand side participation

- Figures compare the forecast median with the observed median
- Legitimate comparison, but potentially not very informative
- More informative to compare the forecast distribution against the observed distribution
- For example, could use side-by-side box plots or violin plots

Figure 65 Evaluation of actual compared to forecast price-driven DSP in New South Wales



Recommendation 22: Consider reporting the accuracy of the demand side participation forecasts by comparing the forecast *distribution* with the observed *distribution*, this could be achieved using side-by-side box plots or violin plots, for example

THE UNIVERSITY of ADELAIDE

Supply – DSP during reliability events

- Interesting section and potentially one of significance
- Presentation is a little confusing
- Descriptions are based on:
  - regional maximum demand days
  - reliability events
- Underlying assumption that maximum demand, high prices and reliability events are highly correlated
- Discussions in the Annual Reports seem to suggest that this correlation may not be as high as one might expect
- Clarity on the purpose and meaning of this section would be useful

Recommendation 23: Investigate ways to clarify the presentation in the “DSP response during reliability events” section to assist the reader’s understanding

THE UNIVERSITY of ADELAIDE

Methodological Recommendations

Methodology

- Recommendations 6 and 21 have a methodological component

Analysis of consumption by industry segments

- The refinement of industry segmentation for consumption forecasting is likely to lead to real improvements in the forecast and assessment of its accuracy
- Investments in this area now will position AEMO well for the future as the economy transitions more rapidly to a low-carbon economy

Recommendation 24: Consider the introduction of further industry segmentation to improve consumption forecasting

THE UNIVERSITY of ADELAIDE

Methodology – weather years

- Weather years from the 2010-11 financial year onwards used in forecasts
- Forecasts can only consider a relatively narrow collection of weather possibilities --- underestimate variance

- Bureau of Meteorology has excellent weather data for the past 50+ years
- Concerns about using weather years from further back in history --- introduce significant bias

- Choice of range of weather years is a variance/bias trade-off

- Synthetic years allow the models to explore a broader collection of **possible** weather outcomes
- Forecasts benefit from being able to use many synthetic weather years without being exposed to significant bias

Recommendation 25: To increase the robustness of the forecast process, increase the number of weather years that are incorporated. This could be achieved by using more historical data or, preferably, by making use of synthetic weather years

THE UNIVERSITY of ADELAIDE

Methodology – voluntary load reductions

- Only plays a small role in the forecasts
  - reported only once in the most recent three years across the five regions

- Scale factors are point estimates with no information regarding their accuracy or precision

- If this feature is regarded as being of reasonable importance, then it is appropriate to investigate the accuracy of the assumed scale factors and whether the forecasts should account for variability

Recommendation 26: Consider upgrading the accuracy and level of assurance of the assumed scale factors in the Potential adjustment – voluntary load reductions feature

THE UNIVERSITY of ADELAIDE

Conclusion

Next steps

- AEMO and stakeholders to provide feedback within the next week
  - please email any feedback to the energy.forecasting@aemo.com.au
- This feedback will be incorporated in the final report, due mid-September

THE UNIVERSITY of ADELAIDE

PRV12105 CRICOS 00123M

make history.

THE UNIVERSITY of ADELAIDE

This item may include material that has been copied and communicated under the Statutory Licence pursuant to s113P of the Copyright Act 1968 for the educational purposes of the University of Adelaide. Any further copying or communication of this material may be the subject of copyright protection.