



Draft Forecast Accuracy Report methodology

July 2020

National Electricity Market

VERSION CONTROL

| Version | Release date | Changes |
|----------------|---------------------|----------------|
| 1 | 27/4/2020 | First version |
| 2 | 14/7/2020 | Second version |

Contents

| | | |
|-----------|--------------------------------------|-----------|
| 1. | Introduction | 5 |
| 1.1 | Rules requirements | 5 |
| 1.2 | Forecasting framework | 5 |
| 1.3 | Related documents | 7 |
| 1.4 | Structure of document | 7 |
| 2. | Adjusting demand | 8 |
| 2.1 | Actual demand | 8 |
| 2.2 | Adjusted demand | 9 |
| 3. | Categories of forecast in use | 14 |
| 3.1 | Representing uncertainty | 14 |
| 3.2 | Deterministic (point) forecasts | 14 |
| 3.3 | Probabilistic forecasts | 15 |
| 3.4 | Summary of forecast categories | 15 |
| 4. | Input forecasts | 17 |
| 5. | Demand forecasts | 18 |
| 5.1 | Operational energy consumption | 18 |
| 5.2 | Maximum and minimum demand | 19 |
| 6. | Supply forecasts | 24 |
| 6.1 | Demand side participation | 24 |
| 6.2 | Supply availability | 25 |

Tables

| | | |
|---------|--|----|
| Table 1 | Example calculation of the impact on demand in MW of customers affected by an outage | 10 |
| Table 2 | Typical voluntary appliance consumption reduction settings | 11 |
| Table 3 | Forecast and actual residential connections growth rate comparison. 2018-19 (%) | 15 |
| Table 4 | Matrix of forecast categories | 16 |
| Table 5 | Example: South Australia energy consumption forecast accuracy by component | 18 |

Figures

| | | |
|----------|---|----|
| Figure 1 | End to end high-level overview of the reliability forecast process | 6 |
| Figure 2 | Relationship between metered demand, actual demand, and adjusted demand | 8 |
| Figure 3 | Example DSP response calculation based on half-hourly load measured across large customers | 9 |
| Figure 4 | Example voluntary demand reduction curve for New South Wales | 12 |
| Figure 5 | Estimated maximum daily residential demand vs maximum daily temperature, New South Wales | 13 |
| Figure 6 | South Australia simulated temperature at time of maximum demand | 15 |
| Figure 7 | Example: Energy consumption component variance chart, South Australia | 19 |
| Figure 8 | Example comparison of forecast price-driven DSP response vs actual for different price triggers | 25 |
| Figure 9 | Example supply availability curve – New South Wales black coal, top 10 hottest days of 2019 | 27 |

1. Introduction

AEMO develops a number of demand and supply forecasts for both operational purposes (short-term forecasts) and reliability/planning purposes (long-term forecasts) for the National Electricity Market (NEM). This document focuses on longer-term electricity demand and supply forecasts and the methodology used by AEMO to assess the accuracy of these.

Within AEMO, these forecasts are used in a number of reliability and planning processes, including Medium Term Projected Assessment of System Adequacy (MT PASA), the Electricity Statement of Opportunities (ESOO) and the associated reliability forecast¹ used for the Retailer Reliability Obligation (RRO), and the Integrated System Plan (ISP). Demand forecasts are also used by industry participants and governments for their own work.

To ensure the insights and advice derived from the forecasts are as accurate as can be expected, AEMO uses a continuous improvement process which includes the assessment of forecast accuracy, determining causes of forecasts deviating from actuals/observed values and identifying and implementing improvements to enhance the forecasts in future years.

The introduction of the reliability forecast under the RRO rules in 2019 increased the importance of the forecast accuracy. To assess if the methodologies applied were fit for purpose, AEMO commissioned an external review of its forecast accuracy assessment methodology undertaken by University of Adelaide. Recommendations arising from the review were adopted by AEMO where practicable to increase the depth and breadth of the its forecast accuracy reporting².

1.1 Rules requirements

AEMO is required to publish an assessment of forecast accuracy at least annually in accordance with the National Electricity Rules (NER) clause 3.13.3A(h):

AEMO must, no less than annually, prepare and publish on its website information on:

- (1) the accuracy to date of the demand and supply forecasts, and any other inputs determined by AEMO to be material to *reliability forecasts*; and
- (2) any improvements made by AEMO or other relevant parties to the forecasting process that will apply to the next *statement of opportunities*,

in accordance with the Reliability Forecast Guidelines (as applicable).

Where availability of information makes comparisons to older statement of opportunities necessary, AEMO may include the statement of opportunities for the preceding 24 months.

While the clause specifically references accuracy of forecasts and other inputs that materially impact the reliability forecast, AEMO's other reliability and planning processes generally share forecasts to ensure consistency, so the forecast accuracy assessment would be equally relevant for all these processes.

1.2 Forecasting framework

The process for producing a full reliability forecast can be split into three overall components:

- Demand forecasts – the forecast load to be met for the NEM.

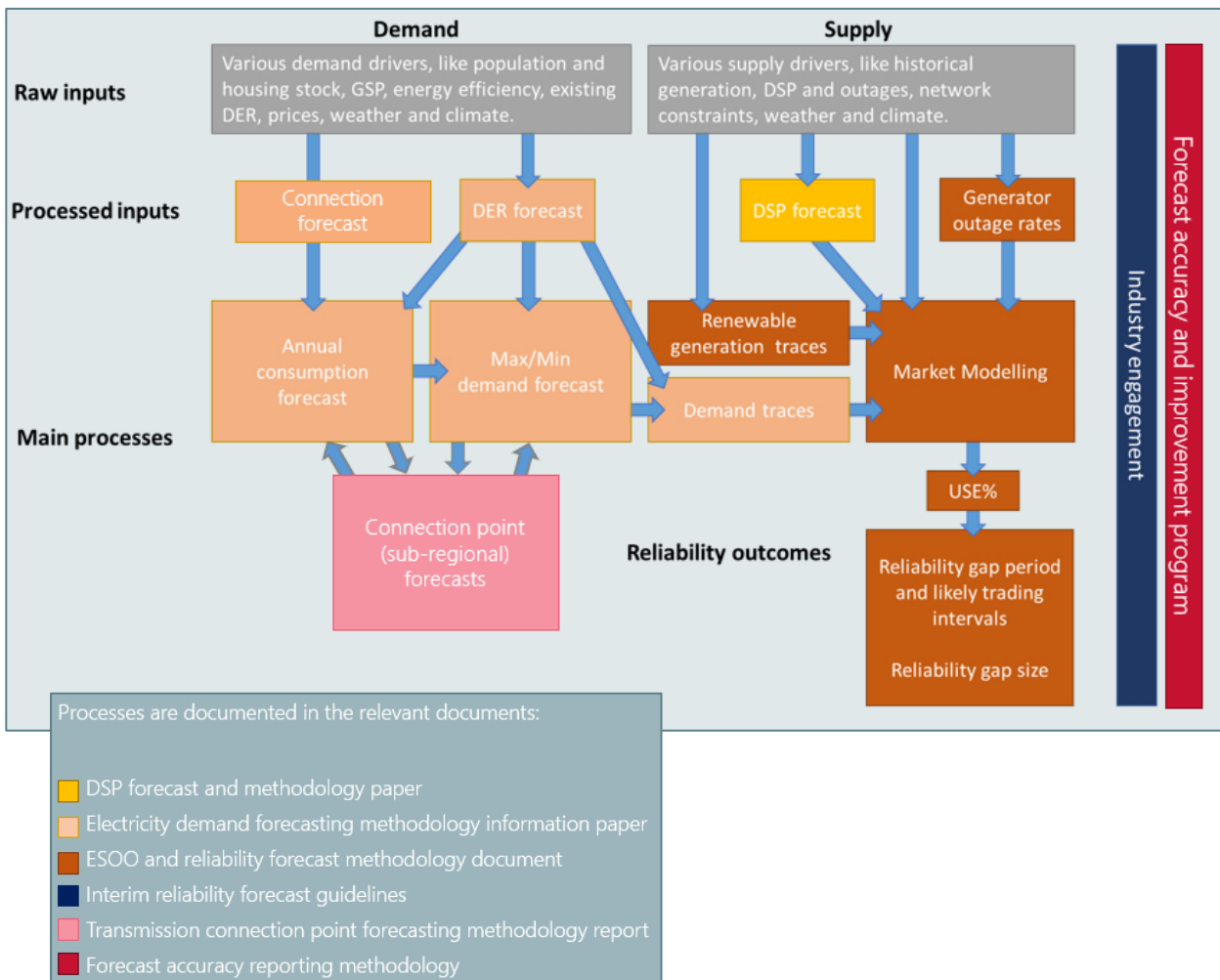
¹ See <https://www.aemo.com.au/energy-systems/electricity/national-electricity-market-nem/nem-forecasting-and-planning/forecasting-and-reliability/nem-electricity-statement-of-opportunities-esoo>.

² See https://www.aemo.com.au/-/media/files/electricity/nem/planning_and_forecasting/accuracy-report/forecastmetricsassessment_uoa-aemo.pdf.

- Supply forecasts – the operational parameters applied for generators, demand side participation (DSP), large-scale storage, and transmission network elements.
- Reliability forecast – the assessment of the ability of available supply to meet demand.

Each of these comprises various components and needs different inputs. Figure 1 provides an overview of the end-to-end process and highlights the different methodology documents that explain the different processes and their inputs. In accordance with the Australian Energy Regulator’s (AER’s) Interim Forecasting Best Practice Guidelines³, fundamental methodologies needed in the forecasting process must be determined using Forecasting Best Practice Consultation Procedures⁴ at least every four years.

Figure 1 End to end high-level overview of the reliability forecast process



The Interim Forecasting Best Practice Guideline also specifies minimum performance analysis requirements for inclusion. In the Interim Reliability Forecast Guidelines⁵, AEMO has committed to meet those requirements in the Forecast Accuracy Report, by including:

- I. an examination of the performance of each forecast component, per NEM region, including:

³ See <https://www.aer.gov.au/retail-markets/retail-guidelines-reviews/retailer-reliability-obligation-interim-forecasting-best-practice-guideline>.

⁴ Available in Appendix A of the Forecast Best Practice Guidelines, link in reference 3.

⁵ See https://www.aemo.com.au/-/media/Files/Stakeholder_Consultation/Consultations/NEM-Consultations/2019/Interim-reliability-forecast-guidelines/Interim-Reliability-Forecast-Guidelines.pdf.

- (A) input drivers of demand;
 - (B) energy consumption (annual assessment);
 - (C) maximum and minimum demand;
 - (D) input drivers of supply;
 - (E) supply availability; and
 - (F) reliability.
- II. an explanation of the results and any material deviation of trend in differences; and
- III. a list of actions undertaken, or to be undertaken, to improve the accuracy of the forecast and forecast components as part of AEMO's forecasting improvement plan.

AEMO's other longer-term demand and supply forecasts, for example the ones used for the ISP, use similar components and will be covered by the discussion of the accuracy of the reliability forecast as well.

1.3 Related documents

In its Interim Reliability Forecast Guidelines⁶, AEMO set out its overall framework for producing its reliability forecast, including its stakeholder engagement. For the detailed methodologies used in producing the forecast, it referred to separate methodology reports:

- Demand Side Participation Forecast and Methodology Paper.
- Electricity Demand Forecasting Information Paper.
- ESOO and Reliability Forecast Methodology Document.

This document completes the suite, explaining how forecast performance is assessed. The latest versions of these documents are available from AEMO's website⁷.

The outcome of the University of Adelaide review of AEMO's forecast performance analysis is also available on AEMO's website⁸.

1.4 Structure of document

This document is structured the following way:

- Section 2 discusses methodologies for adjusting demand observed to match the definition forecast.
- Section 3 identifies four categories of forecast used by AEMO that require differing approaches for assessing accuracy.
- Section 4 discusses methodologies relevant to the forecast inputs.
- Section 5 discusses methodologies relevant to the demand forecasts.
- Section 6 discusses methodologies relevant to the supply forecasts.

⁶ See <https://aemo.com.au/en/consultations/current-and-closed-consultations/interim-reliability-forecast-guidelines>.

⁷ See <https://aemo.com.au/energy-systems/electricity/national-electricity-market-nem/nem-forecasting-and-planning/scenarios-inputs-assumptions-methodologies-and-guidelines>.

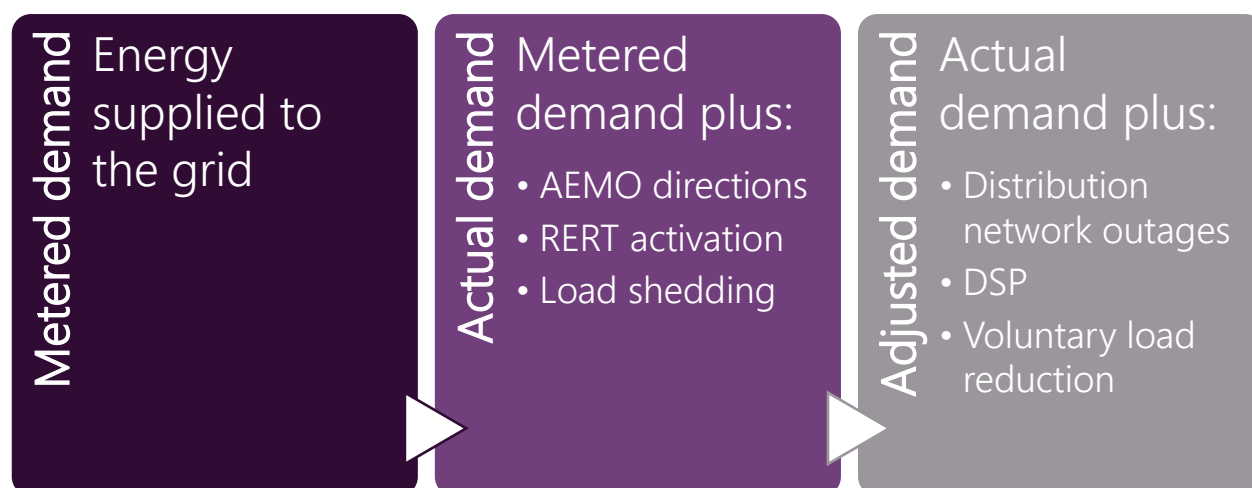
⁸ See https://aemo.com.au/-/media/files/electricity/nem/planning_and_forecasting/accuracy-report/forecastmetricsassessment_uoa-aemo.pdf.

2. Adjusting demand

AEMO's demand forecast represents demand in the absence of any load shedding, use of non-market reserves (RERT), and DSP. DSP is forecast separately, and included in AEMO reliability modelling on the supply side of the equation as per Figure 1. Over the course of the year, such impacts are rare and generally insignificant when comparing actual annual energy consumed with forecast. But on individual maximum and minimum demand days, the impact can be significant and appropriate adjustments are required to be able to compare demand on these days with the Probability of Exceedance (POE) forecast, for example, when assessing the accuracy of these forecasts.

In addition to the metered demand, AEMO operates with two different demand adjustments, as shown in Figure 2.

Figure 2 Relationship between metered demand, actual demand, and adjusted demand



AEMO may apply adjustments for:

- Very high demand days both during summer and winter
- Very low demand days across the year.
- Exceptional, long duration outages of major loads, if affecting annual energy consumption that year.

2.1 Actual demand

Actual demand is defined to be consistent with the RRO as outlined in the Interim Reliability Forecast Guidelines⁹ (Section 6.3) in order to meet the requirements in NER clause 4A.A.4(b). Here, *actual demand* represents metered demand plus the following adjustments:

- Directions by AEMO to generation or loads.
- RERT activated or dispatched by AEMO.
- Load shedding directed by AEMO.

⁹ See https://aemo.com.au/-/media/files/stakeholder_consultation/consultations/nem-consultations/2019/interim-reliability-forecast-guidelines/interim-reliability-forecast-guidelines.pdf.

All these adjustments are controlled by AEMO, allowing *actual demand* to be published soon after the actual event¹⁰. Publishing soon after the actual event means all the components above are based on the amount directed/dispatched/activated by AEMO, rather than a post-event assessment of the actual delivery of these amounts, which (due to the settlement process) may take weeks. So, if AEMO activated 40 MW of RERT contracts and directed 20 MW of load shedding, the adjustment would be 60 MW in total, even though later settlement data may show that the RERT only delivered 38 MW and 25 MW of load was estimated to be impacted by the shedding.

2.2 Adjusted demand

It is appropriate to make further adjustments of demand, beyond those included in the *actual demand* definition above, to estimate what demand would have been under normal circumstances and allow a like-for-like comparison with the forecasts.

AEMO has split the adjustments into two broad categories¹¹:

- **Firm** – these are possible to estimate based on settlement metering data (of individual loads and non-scheduled generators), and cover components like DSP and impacts of distribution network outages. This can also include a more comprehensive estimate of the actual response from activated RERT and AEMO directed loads, non-scheduled generation, and load shedding, if evidence indicates this is substantially different from the directed amount included in *actual demand*.
- **Potential** – these adjustments are more approximate, and are based on an expectation of behavioural responses that cannot be verified (easily) by meter data analysis. This covers cases where the public in advance is asked to conserve energy because the system is forecast to be strained in the coming day.

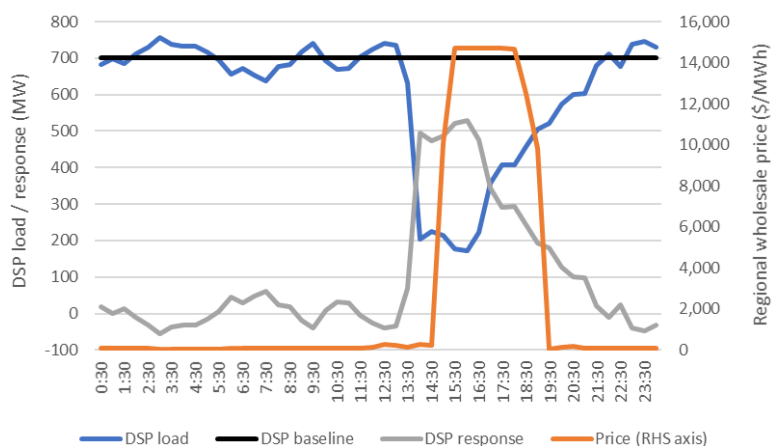
The methodologies used to estimate these adjustments are discussed below.

Firm adjustment – DSP price response

For regional high demand days with high prices (at least one half-hour with wholesale prices exceeding \$1,000/MWh), AEMO estimates DSP response from looking at the metered consumption from all larger customers in the region during the event, and compares this with the period just before and after the event, to define baseline consumption.

Figure 3 illustrates this showing an example day. As the firm adjustment, AEMO will use DSP response for the period (in the example 13:30 to 21:00) where it differs significantly from zero.

Figure 3 Example DSP response calculation based on half-hourly load measured across large customers



¹⁰ Publishing of adjusted demand is still in development, with no firm delivery date.

¹¹ As first presented in Appendix A1 in AEMO's Summer 2019 Forecast Accuracy Update, at https://www.aemo.com.au/-/media/Files/Electricity/NEM/Planning_and_Forecasting/Accuracy-Report/2019-Summer-Forecast-Accuracy-update.pdf.

This approach may leave out DSP from smaller customers. Some of these would be covered under RERT (see Section 2.1 above) or the DSP reliability programs (see below), but overall the combined DSP response used in the firm adjustment is most likely slightly lower than what was actually delivered across all customers.

While AEMO is currently focusing on the larger sites, AEMO will consider the addition of smaller, more varying loads if good estimates can be found for individual days through methodology improvements made to AEMO’s DSP forecasting work.

Firm adjustment – DSP reliability programs

The price response is supplemented by estimated impacts of network reliability programs, as provided by the operating network service provider (NSP). For example, AEMO obtains an estimated response from Ausnet’s Critical Peak Day program after the days where this has been called. AEMO reviews DSP response estimates for materiality and feasibility. The response estimate is also validated against known ranges of response levels for the relevant technology.

AEMO may over time replace the NSP’s estimate with its own, developed from information reported through the DSP Information (DSPI) Portal.

Firm adjustment – distribution outages

AEMO adjusts for any significant outages (events affecting at least 20,000 customers during one of the annual peak demand days). For example, lost customer load due to major storm related outages in the distribution network. Adjustments will be made for all half hours during these events including half hours where less than 20,000 customers were disconnected.

For this adjustment, AEMO seeks an estimate of customers without power for the relevant period from the relevant NSP, or an estimate of the impact in MW directly if available.

If AEMO can only get an estimate of customers not supplied, this is translated into a MW impact using an assumed diversified customer demand of 2 kilowatts (kW) per customer, which reasonably reflects average customer load at time of maximum demand conditions. The 2 kW/customer figure was obtained through a June 2020 survey of NSPs targeting average residential load at time of maximum demand (over a large geographical area); 2 kW was the average of the survey results.

AEMO will not make adjustments for smaller outages (where the maximum number of affected customers is fewer than 20,000), or for outages in general on lower demand days, unless this caused a minimum demand event.

Table 1 below show an example of each estimation rule.

Table 1 Example calculation of the impact on demand in MW of customers affected by an outage

| Date and NEM time (period ending) | Estimate of customers not served | Estimated MW impact |
|-----------------------------------|----------------------------------|---------------------|
| 3 February 2020, 18:30 | 15,000 | 30 MW |
| 3 February 2020, 19:00 | 25,000 | 50 MW |
| 3 February 2020, 19:30 | 7,500 | 15 MW |
| 3 February 2020, 20:00 | 0 | 0 MW |
| 4 February 2020, 17:00 | 6,200 | 0 MW |
| 4 February 2020, 17:30 | 4,000 | 0 MW |
| 4 February 2020, 18:00 | 2000 | 0 MW |
| 4 February 2020, 18:30 | 1000 | 0 MW |

Firm adjustment – re-estimate of directions or RERT activation

Sometimes actual response differs from the AEMO directions or dispatch orders. Once settlement data is available, AEMO may revisit the estimated response if it is found likely it may differ from the directions. In that case, the adjustment will be calculated similar to the DSP price response discussed above. Note: as this will happen weeks after the event, it will only be relevant for assessing forecast accuracy, and will not be used to reassess actual demand for purpose of the RRO.

Potential adjustment – voluntary load reductions

On occasion, state governments or utilities make public appeals (through television, radio, and other media) for electricity users to conserve electricity usage, when possible and safe to do so.

This option to reduce demand is only available if reliability issues have been foreseen the day ahead to allow sufficient time for the message to be disseminated to the public. Most often this is not the case, as issues arise due to sudden compounding impacts of generator and/or transmission outages on high demand days that had not otherwise been seen as extreme. For this reason, unlike DSP, voluntary load reductions of this kind are not included in the ESOO modelling.

AEMO's approach to estimating voluntary load reduction is to utilise the Demand Reduction Calculator (DRC) tool to determine the expected value for each appliance type's total load (in the absence of the voluntary load reduction), and then apply subsequent layers of real-world conditions to arrive at an estimated reduction.

The DRC is primarily used by jurisdictions for emergency planning. It assesses the impact on demand for a mandated reduction in use of different appliance types and is based on estimated daily load profiles per appliance type for each NEM region. While the load profiles are rebased every year, the underlying data on appliances are based on data from the Residential Baseline Study¹², most recently updated in 2015. As the DRC holds the 'expected values' (statistical terminology) of appliance type load, they already incorporate the average usage rate for a particular appliance type at a particular time of day. AEMO uses the DRC, but with a much lower compliance rate to reflect only a fraction of those receiving the message choosing to act.

Thus, AEMO takes the DRC's expected load as a 100% baseline, and from this:

- Removes the proportion of people who don't receive the message promoting voluntary reduction. Typically, AEMO will assume 50% for this factor.
- Then removes the proportion of people who are exempt from responding to the reduction (such as life support equipment). Typically, AEMO will assume 5% for this factor.
- Then applies the voluntary response level (reduction or deferral), as detailed below in Table 2.

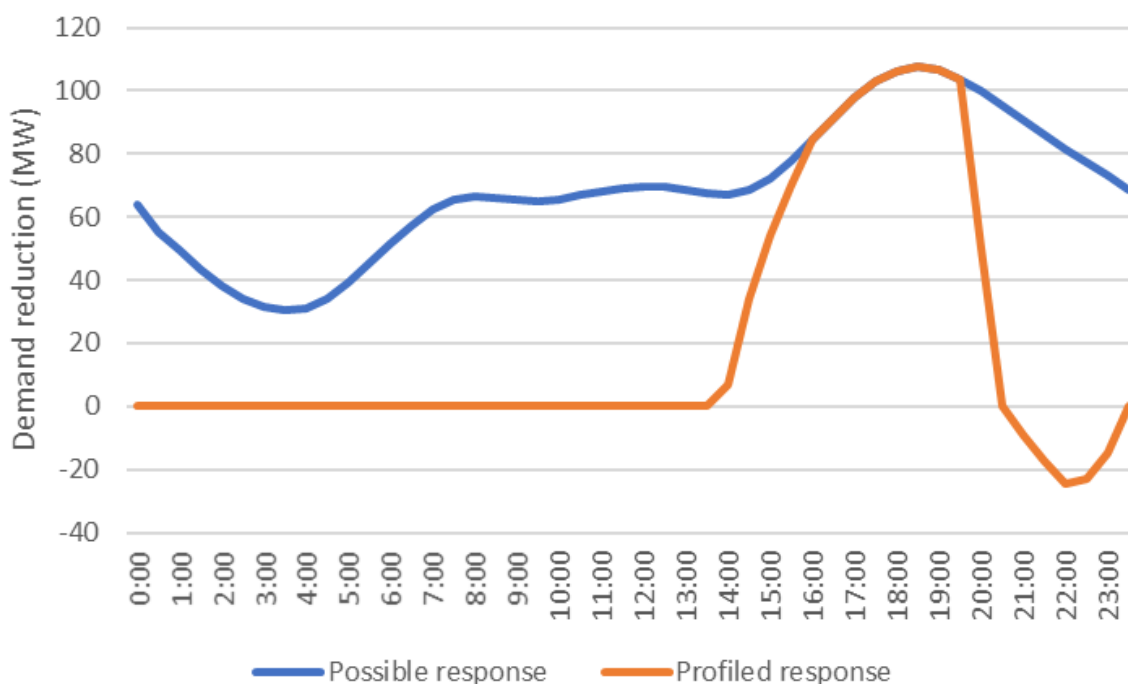
Table 2 Typical voluntary appliance consumption reduction settings

| Appliance type | Assumed reduction in consumption |
|--------------------|----------------------------------|
| Lighting | 10% |
| Pool pumps | 20% |
| Washers/dryers | 50% |
| Dishwashers | 50% |
| Computers and IT | 10% |
| Home entertainment | 10% |

¹² See <https://www.energyrating.gov.au/document/report-residential-baseline-study-australia-2000-2030>.

For a high demand day in New South Wales, this will result in a possible reduction as shown in Figure 4 below.

Figure 4 Example voluntary demand reduction curve for New South Wales



The possible response (blue line) is computed from the weighted average appliance usage rates at the relevant day type and time, and adjusted for:

- Only 50% of customers receiving the energy conservation request.
- 5% with exemptions.
- The assumed voluntary response rates.

The profiled response (orange line) reflects the ramp up from zero response to the full possible response over two hours, and ramp down later, ending up negative (representing increased demand from consumption for appliances that have been postponed, such as dishwashers and washing machines).

Another potential source of voluntary response is from heating/cooling. Typically, reduction requests come on hot days and encourage consumers to set a higher than usual temperature on air-conditioners to lower consumption. Air-conditioner usage is the key driver behind the peak demand days in summer and the potential for reducing demand is significant.

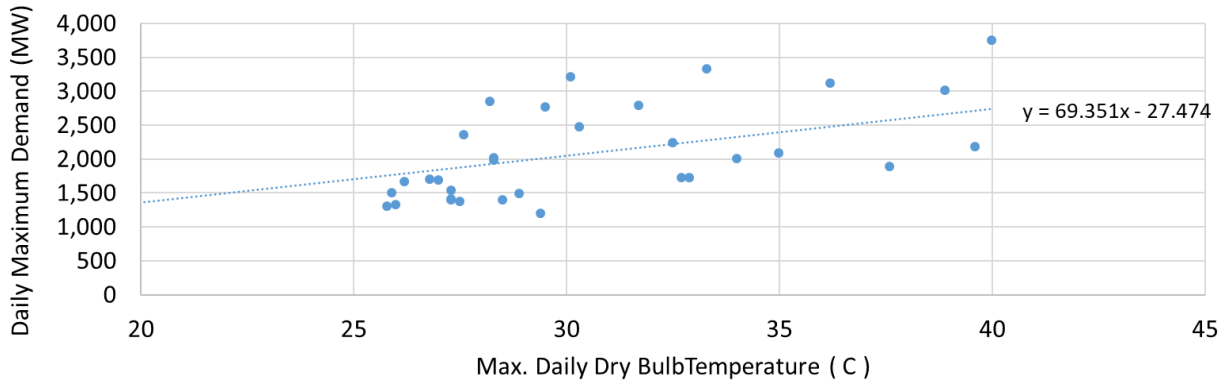
AEMO is still building understanding of the impact of changing the thermostat settings, and currently applies a 50 MW reduction in New South Wales (less in the other regions), using a profile similar to Figure 4, although excluding the rebound at the end.

Analysis of the model used to assess saturation of energy efficiency measures on extreme demand days reveals a three-degree difference in ambient temperature can cause approximately a 200 MW difference in residential cooling load in New South Wales (~70 MW per degree, as per Figure 5). From this, AEMO similarly assumes a three-degree increase in average customer thermostat settings can give a reduction of around 200 MW for the same ambient temperature. Presuming that 50% of all customers get notified, and 50% of these make the adjustment to set the thermostat three degrees higher, the resulting impact is 50 MW.

The 50 MW, combined with the other voluntary response discussed above (160 MW in total), is broadly consistent with previous estimates of the impact of voluntary calls for reduction. For example, for a call for

voluntary reduction of consumption on 10 February 2017, AEMO estimated a 200 MW impact in New South Wales¹³.

Figure 5 Estimated maximum daily residential demand vs maximum daily temperature, New South Wales



¹³ See <https://www.aemo.com.au/-/media/Files/Electricity/NEM/Market Notices and Events/Power System Incident Reports/2017/Incident-report-NSW-10-February-2017.pdf>.

3. Categories of forecast in use

3.1 Representing uncertainty

There are numerous uncertainties that must be considered when forecasting the future of the power system. The nature of the uncertainty varies, depending on the forecast timeframe. For example, a month ahead forecast will predominantly consider uncertainties in consumer behaviour, industrial production, and generator availability; a 10 year ahead forecast will consider broader uncertainties in the economic, social, and technological progression of society.

AEMO represents these uncertainties in two parts:

- Structural drivers, which are modelled as scenarios, including considerations such as:
 - Population.
 - Economic growth.
 - Electricity price.
 - Technology adoption.
 - Generation production and construction costs.
 - Greenhouse gas emission policies.
- Random drivers, which are modelled as a probability distribution, including considerations such as:
 - Weather-driven coincident customer behaviour.
 - Non-weather-driven coincident customer behaviour.
 - Weather-driven generation output.
 - Transmission failure rates and available capacity.
 - Generator failure rates and available capacity.

The above structure used to represent uncertainty applies to both input forecasts (photovoltaic [PV] uptake, generator failure rates), and component-based output forecasts (energy consumption, system reliability).

3.2 Deterministic (point) forecasts

The scenarios are constructed using deterministic forecasts of the structural drivers, meaning each scenario is assigned a set of parameters that describe a future state. These deterministic parameters are not subject to further uncertainty.

The following table is a demonstration of a deterministic input taken from the 2019 Forecast Accuracy Report¹⁴. Each scenario (Slow Change, Neutral, Fast Change) was assigned a forecast rate of residential connections growth. While actual growth could plausibly have taken any value between 0% and 3%p.a., the uncertainty was simplified to three scenarios for easy consumption and comparison.

¹⁴ See https://www.aemo.com.au/-/media/files/electricity/nem/planning_and_forecasting/accuracy-report/forecast_accuracy_report_2019.pdf.

Table 3 Forecast and actual residential connections growth rate comparison. 2018-19 (%)

| | NSW | QLD | SA | TAS | VIC |
|----------------------|------|------|------|------|------|
| Actual (Jun18-Jun19) | 1.4% | 1.4% | 1.0% | 1.2% | 1.9% |
| Slow Change scenario | 1.6% | 1.5% | 1.2% | 0.7% | 1.7% |
| Neutral scenario | 1.8% | 1.7% | 1.3% | 0.8% | 1.9% |
| Fast Change scenario | 2.0% | 1.9% | 1.5% | 0.9% | 2.1% |

Note. The difference between forecast and actual performance in connections growth was the catalyst for a forecast improvement item. See the 2019 Forecast Accuracy Report for more detail.

3.3 Probabilistic forecasts

Random drivers are included in numerous forecasts, through the inclusion of a full probability distribution. Figure 6 demonstrates how large uncertainties in both weather and the consumer response to weather result in a probability distribution of temperatures at time of maximum demand. These random drivers are not subject to any discrete sampling, and forecast outcomes are identified through mathematically applied probability functions, or large numbers of Monte Carlo simulations.

Figure 6 South Australia simulated temperature at time of maximum demand



3.4 Summary of forecast categories

Using the above definitions, it is possible to represent all AEMO reliability forecast components in a simple grid, shown in Table 4.

The four quadrants reflect fundamentally different forecast processes that require four different approaches to measuring and reporting on forecast accuracy. These four categories will be referenced throughout the methodology report to describe the unique characteristic requiring consideration.

Table 4 Matrix of forecast categories

| | Deterministic | Probabilistic |
|---|---|--|
| Input forecasts | <ul style="list-style-type: none"> • Economic and population growth • Energy efficiency • DER uptake • New generator connections • Generator available capacity • Transmission failure rates, losses and available capacity • Atmospheric greenhouse gas concentrations and related impacts <p>simple percentage error metrics are most appropriate</p> | <ul style="list-style-type: none"> • Weather and the related time-series impact on consumer behaviour, transmission capacity, generator output and asset failure rates. <p>qualitative description of accuracy may be most appropriate</p> |
| Component-based output forecasts | <ul style="list-style-type: none"> • Operational energy consumption <p>best to assess the contribution of each input to aggregate accuracy</p> | <ul style="list-style-type: none"> • Minimum and maximum demand • Connection point forecasts • Demand and VRE traces • Demand side participation • Supply availability • Reliability <p>challenging to assess accuracy using a single observation, requiring exploratory analysis and qualitative justification</p> |

4. Input forecasts

Electricity demand and supply forecasts are predicated on a wide selection of inputs and assumptions. Models incorporate numerous forecast components, including:

- Economic growth and population.
- Distributed PV and behind-the-meter batteries.
- Energy efficiency and appliance mix.
- Electric vehicles (EVs).
- New generator connections.
- Generator forced outage rates.

Some of these forecasts are provided to AEMO by external consultants, while others are developed internally. This section describes the methods that are used to assess the accuracy of these forecasts.

The purpose of assessing the accuracy of input forecasts is to determine whether the scenario settings of the structural drivers are a good reflection of what happened. Given the importance of the Central/Neutral scenario in reliability analysis, most performance analysis will focus on the accuracy of this scenario, unless there is good reason to explore another.

Most inputs are deterministic, the most notable exception being weather. If actuals are available, are assessed by measuring the percentage difference between actual and forecast values of the published forecasts. There are four methods for calculating percentage error, that may vary the calculated error if used interchangeably.

| Calculation | Discussion |
|---|--|
| $percentage\ error = \frac{actual - forecast}{actual} \times 100$ | A positive number indicates the actual is above forecast. Most statistically accurate with a somewhat intuitive interpretation for many users. |
| $percentage\ error = \frac{actual - forecast}{forecast} \times 100$ | Intuitive for those who think of forecasts as budgets. A positive number indicates the actual is above forecast. Introduces statistical bias when evaluating performance across multiple models. |
| $percentage\ error = \frac{forecast - actual}{actual} \times 100$ | A positive number indicates the forecast was an over-estimate of actual. Statistically accurate but less intuitive for some users. |
| $percentage\ error = \frac{forecast - actual}{forecast} \times 100$ | A positive number indicates the forecast was an over-estimate of actual. Introduces statistical bias when evaluating performance and is less intuitive for many users. |

AEMO uses the third method for all accuracy assessments to ensure consistency, as a trade-off between statistical accuracy and ease of interpretation.

AEMO publishes forecast and observed values alongside forecast accuracy metrics for all forecast components where actual values are available. Values may be published in either graph or tabular format. Where an input is subject to confidentiality requirements, AEMO may choose to either aggregate or not publish updated data.

Given the complex multiple variable nature of weather, the assessment of weather will remain entirely qualitative and descriptive. AEMO includes many weather years in all simulations, so discussion will consider whether the observed weather was materially different from the weather years included, and any other distinguishing features.

5. Demand forecasts

This section discusses the methodology used to assess the accuracy of the demand forecast components in AEMO’s forecast accuracy reporting. The DSP forecast is covered in the Supply section (see Figure 1).

The purpose of assessing accuracy of demand forecasts is to determine any material bias in the forecasts and identify the contribution of the input forecasts to aggregate accuracy. While the percentage error metric shown in Chapter 4 is applicable, it would provide no guidance on sources of inaccuracy. Assessments of accuracy must identify the sources of inaccuracy so proposed improvements can target the inputs or models that will generate the largest improvement in accuracy.

5.1 Operational energy consumption

AEMO forecasts annual energy consumption by region, on a financial year basis, for each pre-defined scenario. Given the relatively low influence of DSP, outages, and load shedding on consumption volumes, operational energy consumption is not subject to adjustments.

To better interrogate the drivers of forecast accuracy, AEMO extends the percentage error metric previously discussed. The operational energy consumption model is built using an econometric regression of the scenario input variables. For each relevant input variable, the forecast, actual and percentage difference is reported; as well as the impact on aggregate accuracy. The coefficient from the econometric model can be used to identify the impact as per the following equation 1 (or similar depending on forecast model specification), while the equation for the waterfall component is expressed in equation 2.

$$1. \text{indicative impact on total generation} = \frac{\text{input coefficient} \cdot (\text{input forecast} - \text{input actual})}{\text{OPGEN actual}} \times 100$$

$$2. \text{error from forecast component} = \text{input coefficient} \cdot (\text{input actual} - \text{input forecast})$$

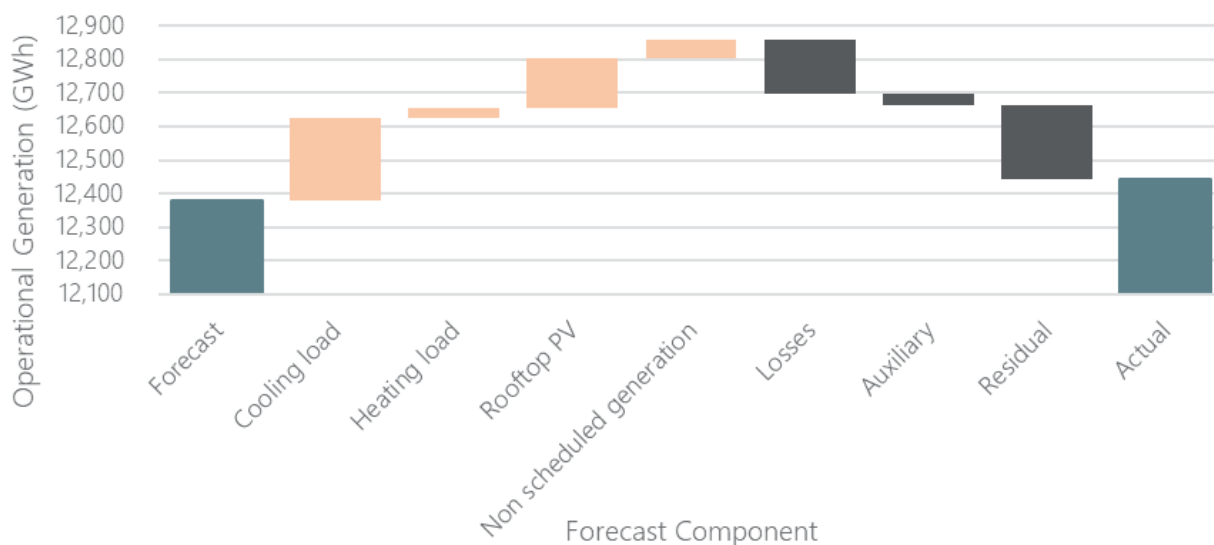
This metric brings context to the input variable inaccuracy, where large variable inaccuracy may only have negligible impacts on total generation. All residuals not explained by input variables will be reported separately, and are likely indicative of model (not input) error. Just like the model specification itself, this method assumes the input variables are independent of each other. The following examples show the South Australia energy consumption component variance table and chart with commentary to demonstrate interpretation.

Table 5 Example: South Australia energy consumption forecast accuracy by component

| Category | 2018 Neutral forecast | Actual | Difference (%) | Indicative impact on total generation |
|--------------------------------------|-----------------------|----------|----------------|---------------------------------------|
| Cooling degree days | 436.3 | 685.1 | -36.3% | -2.0% |
| Heating degree days | 724.6 | 659.3 | +9.9% | -0.2% ¹⁵ |
| Rooftop PV (GWh) | 1,523.0 | 1,373.7 | +10.8% | -1.2% |
| Small non-scheduled generation (GWh) | 228.2 | 173.1 | +31.8% | -0.4% |
| Network losses (GWh) | 1,131.2 | 966.6 | +17.0% | +1.3% |
| Operational sent out (GWh) | 12,053.4 | 12,147.1 | -0.8% | -0.8% |
| Auxiliary load (GWh) | 325.5 | 294.2 | +10.6% | +0.3% |
| Operational as generated (GWh) | 12,378.9 | 12,441.3 | -0.5% | |

¹⁵ Despite reduced heating load for residential customers, heating related business variance was positive and greater.

Figure 7 Example: Energy consumption component variance chart, South Australia



Example interpretation

The annual energy consumption forecast (12,379 gigawatt hours [GWh]) is shown as the first bar in the waterfall chart, while the actual energy consumption for 2018-19 (12,441 GWh) is shown as the last bar, a value that is 0.5% higher than forecast. The bars in between indicate the component contribution to the difference between forecast and actual values, where light orange represents a positive impact, and charcoal represents a negative impact. The residual captures all difference not explainable by the variations in identified input components.

In this example, there are positive contributions from cooling load, heating load, rooftop PV and non-scheduled generation, and a negative contribution from losses and auxiliary. The difference between forecast and actual bars that is attributable to these variables is best resolved through examination of the input forecasts. Forecast difference that cannot be explained by the components and falls into the residual may be caused by input variables that cannot be measured, or by the energy consumption model itself. In this example, the small positive residual implies that, given observed inputs, the forecast model would have expected an actual that was slightly higher and a small downward revision in forecast may be considered alongside input forecast updates.

5.2 Maximum and minimum demand

AEMO produces forecasts of the probability distributions of seasonal minimum and maximum half-hourly demand. These forecasts are compared against adjusted demand, as explained in Section 2, as if no load shedding or demand side participation has occurred. The purpose of the forecast performance assessment is to understand any sources of inaccuracy, so that improvements can target the inputs or models that will generate the largest accuracy improvements.

These forecasts are developed through a computationally intensive simulation process and are summarised for industry as 10%, 50%, and 90% POE forecasts. Given each year comprises a single actual, while the forecast comprises a wide range of possible occurrences, assessing accuracy is challenging. While weather is a large source of the uncertainty represented in the probability distribution, it is not the only driver. Customer demand is highly erratic and only becomes forecastable when aggregated among many customers. While weather drives customers towards coincident appliance use, other factors remain a large variant – the “known unknowns”. Even when all measurable variables are known, there will still be a wide range of possible outcomes driven by the uncertainty in coincident consumer behaviour.

There are numerous methods available to assess the performance of these forecasts. Some of these methods are described below.

Qualitative comparison of observed demand to the forecast distribution

The primary method of reporting accuracy of minimum/maximum demand forecasts is a qualitative comparison, specifying where on the forecast distribution the observed minimum/maximum demand lies and providing contextual factors that may explain this. For example:

“In New South Wales 2017, maximum demand occurred on 10 February 2017, when the temperature reached 43.7°C. The actual maximum demand may have been higher if it hadn’t been for a general call for reduced consumption and engagement of DSP. Accounting for an estimated combined 490 MW of load reductions, the adjusted maximum demand exceeded the forecast of 10% POE demand.”

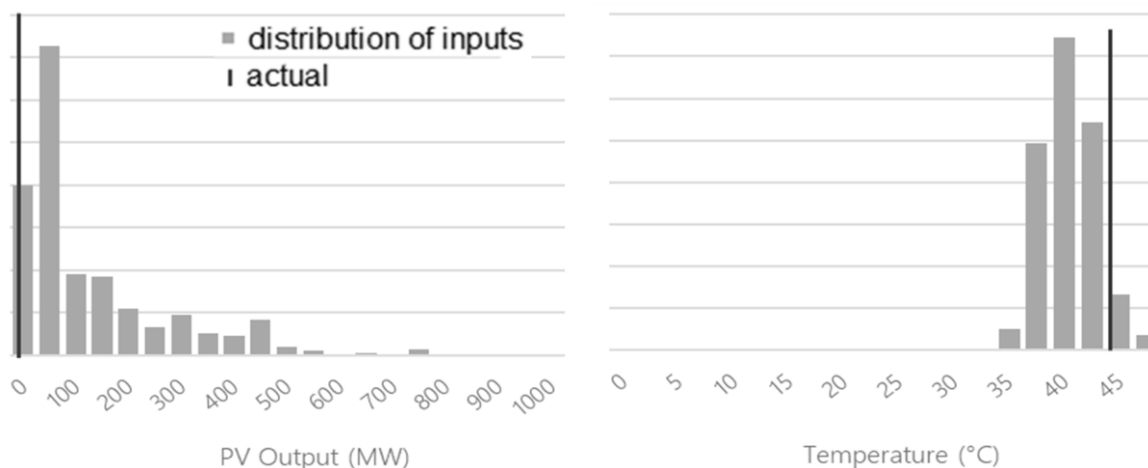
Percentage error of actual relative to the 50% POE forecast

This method follows the same process expressed in Section 4, simply comparing the actual to the 50% POE (median) of the forecast distribution. While easily understood, the accuracy assessment ignores the remainder of the forecast distribution and provides no insights as to the cause of the difference from the 50% POE, or whether the difference is expected. For example:

| 90% POE | 50% POE | 10% POE | Actual | P.E. |
|---------|---------|---------|--------|------|
| 288.03 | 295.63 | 305.50 | 286.99 | +3% |

Comparison of the distribution of individual key drivers

Alongside the forecast distribution of minimum/maximum demand, the distribution of key drivers at the (simulated) minimum/maximum demand(s) can also be reported. Reporting these quantities at the observed minimum/maximum demand interval, compared to the distribution that produced the forecast, provides valuable context. For example:



Example interpretation

The actual summer maximum demand for South Australia was above the 10% POE forecast.

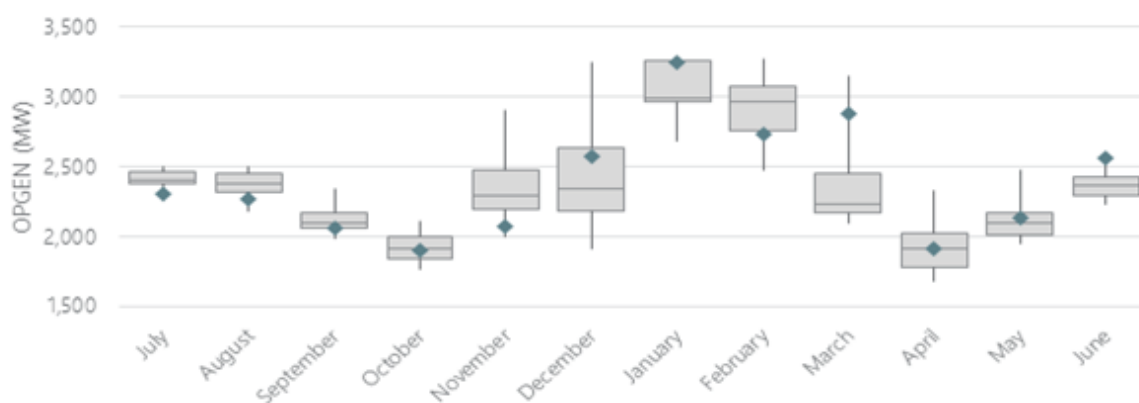
South Australia experienced a hot period in late January, with an annual maximum temperature recorded earlier in the day of 46.6°C. The day also saw the hottest minimum daily temperature of 30.7°C.

The maximum was expected to occur late in the day, when PV output was low or zero. Actual maximum demand occurred late in the day, with zero PV output, at a temperature towards the upper end of the expected temperature distribution.

Monthly demand trace forecast performance

Using operational energy consumption, summer and winter maximums, and minimum demand forecasts, AEMO develops half hourly demand traces that match reference years for market modelling. To further explore demand forecast accuracy, the observed monthly maximum demand can be plotted against a box and whisker plot of the monthly maximums forecast in the demand traces. AEMO uses both 10% and 50% POE traces to develop the box and whiskers plot. These plots provide additional context as any biases or inaccuracies observed in energy forecasts, or extreme demand forecasts will likely also be visible in the monthly outcomes. An example (from the 2018 Forecast Accuracy Report¹⁶) is shown below, where almost all observed monthly maximum values lie within the forecast range.

Figure 33 South Australia monthly maximum demand based on demand traces with actuals



Probabilistic forecasting statistics

There are numerous statistics designed to evaluate the performance of probabilistic forecasts, including the Kolmogorov-Smirnov statistic, Mean Absolute Excess Probability, Score, and Relative Score. These statistics all evaluate the degree to which a forecast distribution matches a set of actuals, but all rely on as large a set of actuals as possible.

In this instance, the forecast is focused on the distribution of possible annual maximum demands, of which there is only one new actual per year. Therefore, the single additional data point each year provides limited value in assessing accuracy out of sample. Additionally, the statistics provide no insight into the source of the inaccuracy, should it be driven by model inputs rather than the model itself. In the recent review of forecast accuracy metrics, the University of Adelaide made numerous recommendations about these statistics for AEMO internal reporting and model evaluation purposes. There is limited benefit for external, particularly non-statistical audiences.

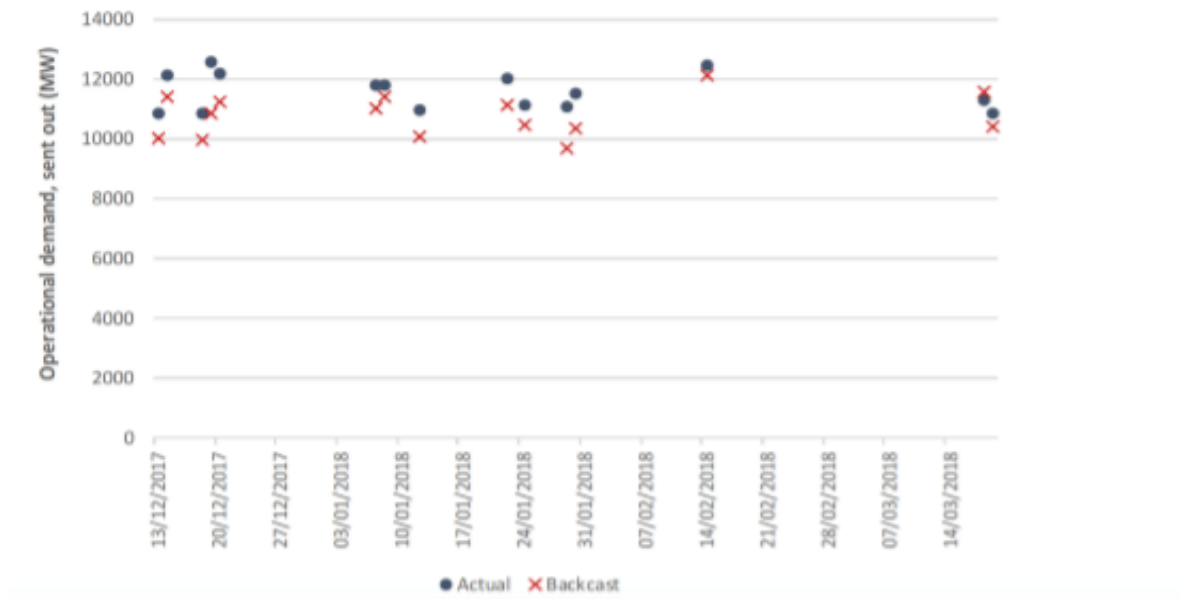
Backcasting

Backcasting has been used by AEMO previously to evaluate the forecast regression fitted to observed inputs, presented without stochastic volatility for the 15 highest observed demand intervals in each region. AEMO demand models, however, utilise stochastic volatility to model elements of coincident consumer behaviour. As such, the backcasts of days with high levels of coincident consumer behaviour always indicate an apparent underforecast, while in reality the backcast does not provide relevant information regarding the accuracy of the forecast. In the recent review of forecast accuracy metrics, the University of Adelaide suggested discontinuation of this metric. An example (from the 2018 Forecast Accuracy Report¹⁷) is shown below, where almost all backcast values lie below observations.

¹⁶ See https://www.aemo.com.au/-/media/Files/Electricity/NEM/Planning_and_Forecasting/Accuracy-Report/Forecast-Accuracy-report-2018.pdf.

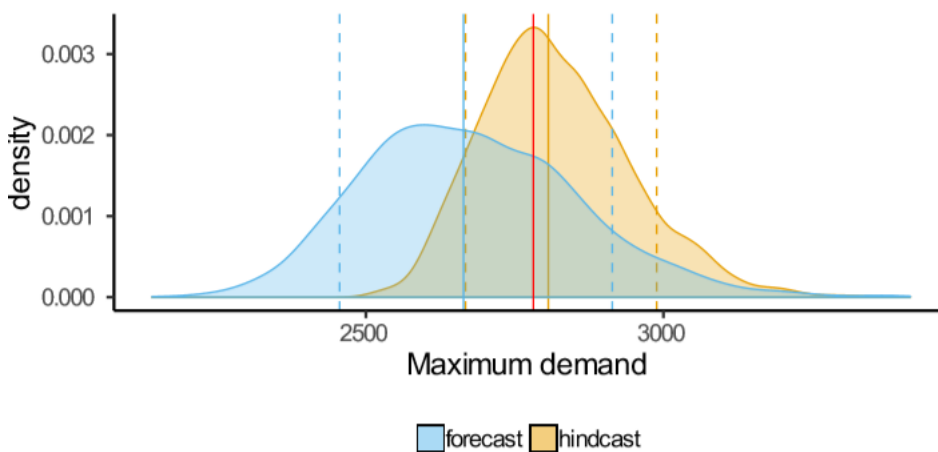
¹⁷ See https://www.aemo.com.au/-/media/Files/Electricity/NEM/Planning_and_Forecasting/Accuracy-Report/Forecast-Accuracy-report-2018.pdf.

Figure 12 Actual versus backcast max demand half-hour for top 15 highest demand days in New South Wales



Hindcasting

Hindcasting involves comparing a forecast that was made historically with a forecast that would be made now for that year, including inputs as actually observed. The purpose of this method is to compare the forecast distribution without known inputs to the forecast distribution using the actual input drivers, thereby elucidating the impact of the forecast drivers on the forecast. The example below shows the difference between the forecast and the hindcast. In this example, given known inputs, the probability distribution has narrowed.



This method describes the scale and direction of the forecast error, and the shape of the unexplainable components, represented by the hindcast distribution width. If the actual falls outside the hindcast, it indicates that the actual was unlikely based on the model, however a single new observation does not confirm model bias, and as it does not interrogate the input variables, it does not indicate the source of any error.

Summary

The methods for describing accuracy of maximum and minimum demand forecasts are summarised below.

| Method | Discussion |
|--|---|
| A qualitative comparison of the observed demand to the forecast distribution | Easy to implement and very useful for establish context about the attributes of the event in question. |
| Percentage error of actual relative to 50% POE | Easy to implement and understand but fails to provide context about the performance of the probabilistic forecast, or any contribution from the input variables. |
| A comparison of the distribution for individual key drivers. | More complex to implement and understand but provides supplementary context about the contribution of the input variables, and any performance issues in the demand models themselves. |
| Monthly demand trace forecast performance | Easy to implement and understand and provides additional context and verification of performance from the array of forecasts that input to the demand traces. |
| Probabilistic forecasting statistics | Particularly challenging to implement with few observations, and to understand. Provide little context about the contribution of input variables to the performance of the forecast. |
| A backcast of the top observed demand intervals | More complex to implement and understand and is inconsistent with the forecast methodology, providing no context about the performance of the forecast. |
| A hindcast of the forecast distribution with inputs as observed | More complex to implement and understand but provides useful insights about the performance of the forecast, once all input variables are accounted for. |

AEMO uses the first, third, and fourth methods in combination for all accuracy assessments to ensure consistency and ease of understanding for a non-statistical audience. In combination, these should sufficiently elucidate the source of forecasting inaccuracies for the effective development of an improvement plan.

AEMO also uses the sixth method when it considers that the significant additional performance analysis is warranted to gain a greater understanding of the probabilistic forecast accuracy.

6. Supply forecasts

This section discusses the methodology used to assess the accuracy of the supply forecast components in AEMO's forecast accuracy reporting.

The purpose of assessing accuracy of supply forecasts is to determine any material bias in the forecasts and identify the contribution of the input forecasts to aggregate accuracy. While the percentage error metric shown in Chapter 4 is applicable, it would provide no guidance on sources of inaccuracy. Assessments of accuracy must identify the sources of inaccuracy so proposed improvements can target the inputs or models that will generate the largest improvement in forecast performance.

6.1 Demand side participation

AEMO forecasts DSP for use in its medium- to long-term reliability assessments (MT PASA, Energy Adequacy Assessment Projection [EAAP], and ESOO) as well as the ISP. It represents reduction in demand from the grid in response to price or reliability signals. In AEMO's modelling, it is treated similarly to supply options as a way to ensure demand can be met.

AEMO forecasts two broad categories of DSP¹⁸:

- Market-driven responses – this category includes residential, commercial, and industrial responses that are typically triggered in response to high electricity prices.
- Reliability event responses – this category includes responses that are called on when power system reliability requires support. They are most common under Lack of Reserve (LOR) conditions, although they often also coincide with high wholesale prices. These responses can be contracted.

Price-driven DSP responses vary significantly from time to time, even for identical market price outcomes, as the decision to respond (or not) is affected by a number of external factors, including contracting levels and duration of the price spike¹⁹. AEMO has therefore developed a process that investigates the recent (three-year) historical distribution of DSP responses for different price triggers, and AEMO has adopted the 50th percentile of historical response as its estimate of expected DSP (half the time response will be higher, half the time it will be lower) for each trigger level.

AEMO's assessment of DSP forecast accuracy is based on two components:

- An assessment of the median (50th percentile) observed DSP response for various wholesale price triggers²⁰ during the most recent year compared to forecast median response.
- An assessment of the estimated DSP response during the regional maximum demand events against the forecast DSP reliability response.

The first component assesses the accuracy of the forecast price response in general when comparing with the estimated actual response. The second assessment is to check if the estimate is reasonable for the highest demand days in particular.

Observed median response by price trigger

The observed median response is found using the same tool and baseline approach (to assess individual DSP response outcomes) as used to produce the DSP forecast. The forecast assesses the response as the median

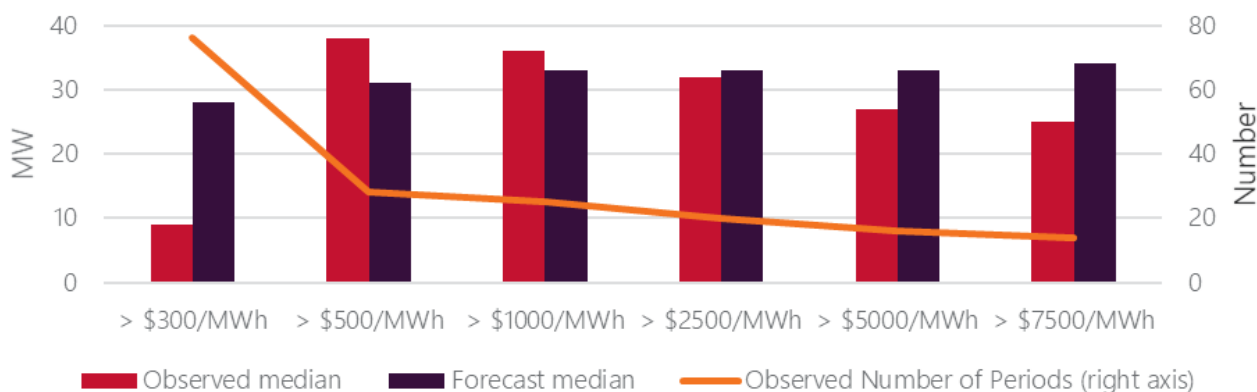
¹⁸ See further details in https://www.aemo.com.au/-/media/Files/Electricity/NEM/Planning_and_Forecasting/NEM_ESOO/2019/Demand-Side-Participation-ForecastMethodology-2019.pdf.

¹⁹ This is discussed in more detail in the Demand Side Participation Forecast Methodology consultation paper from February 2020, at https://www.aemo.com.au/-/media/files/stakeholder_consultation/consultations/nem-consultations/2020/demand-side-participation/demand-side-participation-forecast-methodology-consultation-v4.pdf.

²⁰ AEMO currently models the following price triggers: \$300/MWh, \$500/MWh, \$1000/MWh, \$2500/MWh, \$5000/MWh, and \$7500/MWh.

of all estimated responses each time a wholesale price trigger was reached in the three years leading up to the forecast being produced. This is compared with estimated response in the most recent year, calculated using the same methodology for the actual consumption and price data for the year following the forecast release. In some years, some NEM regions may experience very few high price periods that breach the price triggers used when creating the forecast, in which case it may be impossible to estimate a median response that can be compared meaningfully with the forecast.

Figure 8 Example comparison of forecast price-driven DSP response vs actual for different price triggers



Example interpretation

- Assessing the last year of actual response to various price levels, the median response (of the 76 cases) for regional wholesale prices exceeding \$300/MWh is 9 MW, while for the 14 pricing events above \$7500 is 25 MW. The latter is 8 MW less than forecast, but most other price bands have a closer alignment, and is considered within the accuracy of the forecasting methodology.
- The median response may not be meaningful for very small number of observations. Generally, AEMO will only seek to draw conclusions if the number of pricing events is above 10, which is the case across all pricing levels considered here.

For cases with a reasonable number of price events above the threshold, AEMO will discuss the alignment between forecast and actuals.

Observed DSP during observed regional maximum demand

For the regional maximum demand days, AEMO estimates the price-driven response similarly to the above, though in more detail as responses by individual sites are typically examined and validated rather than looking at the aggregate response only. The analysis is also used to calculate DSP adjustments to historical demand, as discussed in Chapter 2.

For the reliability response, AEMO currently either estimates the response – similar to the way price-driven response is estimated – or asks the participant controlling the program for an estimate. AEMO then discusses how well the observed combined DSP response aligns with the forecast DSP response for the region.

6.2 Supply availability

Generator supply availability is particularly important in reliability studies given it is commonly a key driver of unserved energy (USE) estimates. Supply forecasts are therefore assessed by the degree to which capacity availability estimated in the ESOO matched actual generation availability at times of highest supply scarcity

risk. To achieve this goal, the following method is used to compare ESOO simulations with outcomes during extreme temperature periods in the observed summer.

Extreme temperature periods are likely to align closely with periods of very high demand, possible derating, and possible supply shortfalls. These periods allow exploration of forecast versus actual supply availability considering:

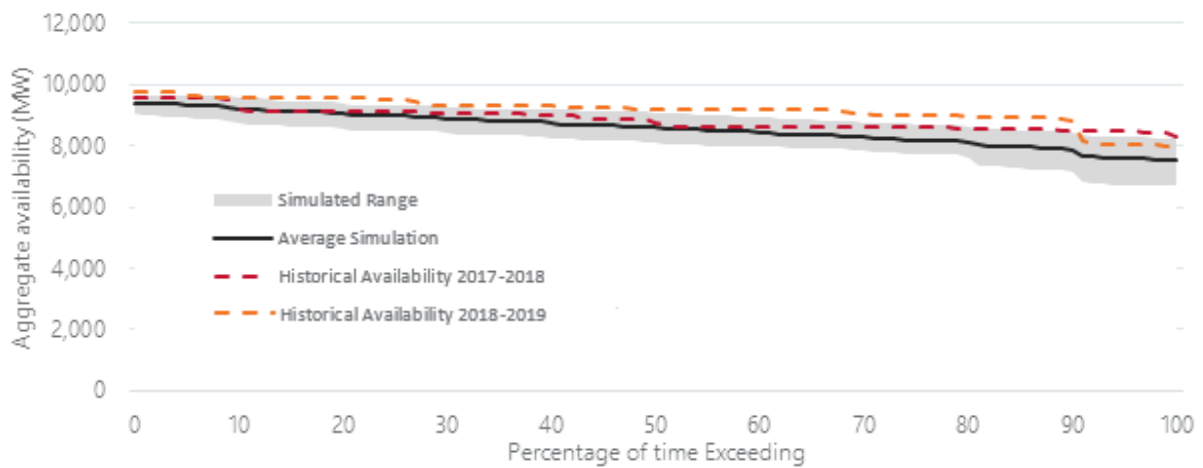
- Available capacity considering de-rating.
- Full unplanned outages.
- Partial unplanned outages.

The method for assessing supply forecast performance involves:

- Selecting availability data for the 10 hottest days observed over summer per region.
 - AEMO will use PASA availability where submitted data quality is high, else will use maximum availability.
 - Eight intervals are chosen per day, including the time of maximum temperature and the seven half-hour periods that followed.
 - This selection of historical data is used to observe generator performance at times of high temperature. High temperature periods are very likely to be linked with periods of tight supply-demand balance, and also represent periods where the physical capability of generator units is most at risk of issues including temperature derating.
 - Units with availability below their listed seasonal availability during these periods are assumed to be experiencing a partial or full outage, rather than a strategic withdrawal of capacity.
- Selecting equivalent forecast availability from ESOO simulations.
 - Simulated availability is taken from 1,000 samples of 10 random days/iterations. The availability data from these days is taken from the maximum temperature period and the seven half-hourly periods that follow (this is to match the number of hours with historical). The 97.5th and 2.5th percentiles of the simulation outcomes are shown to represent the forecast band and eliminate outliers that may occur with very low probabilities.
- Aggregating historical and forecast data for comparison with respect to generation fuel types and regions, plotting duration curves to compare the data sets.
 - Historical trends per fuel type are cleaned such that only units currently operating are considered.

Supply availability curves are presented for all regions, split by variable renewable energy and scheduled generation. Where there is value in exploring specific fuel sources with material contributions to reliability risk, additional availability curves may be added to complement regional analysis. An example of this method is shown below.

Figure 9 Example supply availability curve – New South Wales black coal, top 10 hottest days of 2019



Note. The difference between forecast and actual performance from the fleet of generators was the catalyst for a forecast improvement item. See the 2019 Forecast Accuracy Report for more detail.

Example interpretation

The grey range represents the simulated availability, while the dotted lines represent availability as observed. In this example, actual availability remained within or above simulated availability over the study period, indicating better than expected performance from the fleet of generators. This may be due to cooler than anticipated temperatures or lower than expected failure rates.