Est. 2008

Oakley Greenwood

Phase 2 - Baseline Methodology and Participant Testing

Wholesale Demand Response Mechanism - Baseline Methodology Testing and Metrics

prepared for:
Australian Energy Market Operator

Melbourne +613 9016 2550 **Sydney** +612 8091 3650 **Brisbane** +617 3283 3249

DISCLAIMER

This report was commissioned by the Australian Energy Market Operator (AEMO) to assess the suitability of a selected set of baseline methodologies for use in the Wholesale Demand Response Mechanism. The analysis and information provided in this report is derived in whole or in part from information provided by a range of parties other than Oakley Greenwood (OGW). OGW explicitly disclaims liability for any errors or omissions in that information, or any other aspect of the validity of that information. We also disclaim liability for the use of any information in this report by any party for any purpose other than the intended purpose.

DOCUMENT INFORMATION

Project	Wholesale Demand Response Mechanism – Baseline Methodology Testing and Metrics
Client	Australian Energy Market Operator
Status	Final Report
Report prepared by	Lance Hoch (lhoch@oakleygreenwood.com.au) Angus Rich (Arich@oakleygreenwood.com.au) Alex Cruickshank (acruickshank@oakleygreenwood.com.au) Mark Darlington (mark.darlington@optimate.net.au) Dennis Lilford (dennis.lilford@optimate.net.au)
Date	March 2021

CONTENTS

FIGURES

TABLES

Executive summary

Background

The Baseline Methodology project was undertaken in response to the Wholesale Demand Response Mechanism (WDRM) Rule change. The Rule change establishes a new category of market participant, the Demand Response Service Provider (DRSP). DRSPs will be able to bid demand response into the NEM's wholesale electricity market in direct competition with traditional generators[1]. DRSPs will be able to engage directly with customers – without the involvement of the customer's retailer – for the provision of DR under the WDRM.

The WDRM is expected to provide greater opportunity for consumers to participate in the wholesale electricity market. This is likely to increase competition in that market, and thereby put downward pressure on wholesale electricity prices.

Project purpose

This project was undertaken in two Phases to provide information for use by AEMO in determining the initial baseline(s) to be used in the WDRM (Phase 1) as well as deciding on the approach to be used in eligibility assessment and compliance testing of WDRM participants (Phase 2). This report focuses on Phase 2.

The role of eligibility assessment and compliance testing

AEMO has advised Oakley Greenwood that its current thinking regarding the implementation of eligibility assessment and compliance testing is that:

- Customers meeting the WDRM size threshold will be able to register for participation at any time.
- Eligibility assessment testing will be undertaken when a registration request is received. The eligibility assessment will test the accuracy at which the BM to be used in the WDRM can predict the load of (and therefore the demand response provided by) each NMI seeking to participate in the program. Where eligibility testing indicates the load of the NMI can be predicted by the BM at the level of accuracy set for the program, the NMI will be eligible to participate.
- Compliance testing will be undertaken twice a year, most likely prior to the 1st and 3rd quarters which is when higher prices typically occur in the NEM, making the use of demand response more likely.
- The performance of the baseline methodology itself may be reviewed at AEMO's discretion as needed in order to improve baseline accuracy and/or applicability.

Approach modelled

Several approaches to eligibility assessment were considered. The approach selected in consultation with AEMO for modelling in Phase 2 required each NMI seeking to participate to produce:

1 In doing so, the DR that is bid will be treated essentially as a scheduled load. The scheduled load category has always been a part of the NEM but until now has only been accessible by retailers and registered customers.

- An RRMSE of 0.1 or less in 90% of the days in the time period being used for the eligibility assessment
- For both the 7:00 to 9:00am event window AND the 3:30 to 8:00pm event periods on each relevant day (i.e., weekday or weekend day, excluding public holidays and days of DR activation).
- Over the course of previous calendar quarter.

The modelling of this eligibility assessment approach was run on Large Commercial and Industrial Customer segment (i.e., non-residential customers with annual consumption in the range of 750 MWh to 100 GWh). The modelling used metering data from calendar year 2019 and a 10 of 10 BM with a multiplicative adjustment applied in the pre-event period only and capped at 20%[2].

The rationale for the selecting this approach was as follows:

- The inclusion of both the morning and afternoon/evening windows provides coverage of the vast majority of the half hours in which high prices were found to occur from 2017 through 2019
- The use of a single RRMSE metric was chosen for its simplicity and ability to be easily explained to customers and DRSPs
- The combination of a higher level of accuracy (RRMSE of 0.1 as compared to the RRMSE threshold of 0.2 that is used in the RERT) and a requirement for that level of accuracy to be met on 90% of the days included in the eligibility assessment period was seen as providing an adequate balance between accuracy[3] and the proportion of NMIs that would be able to participate in the WDRM.

Results

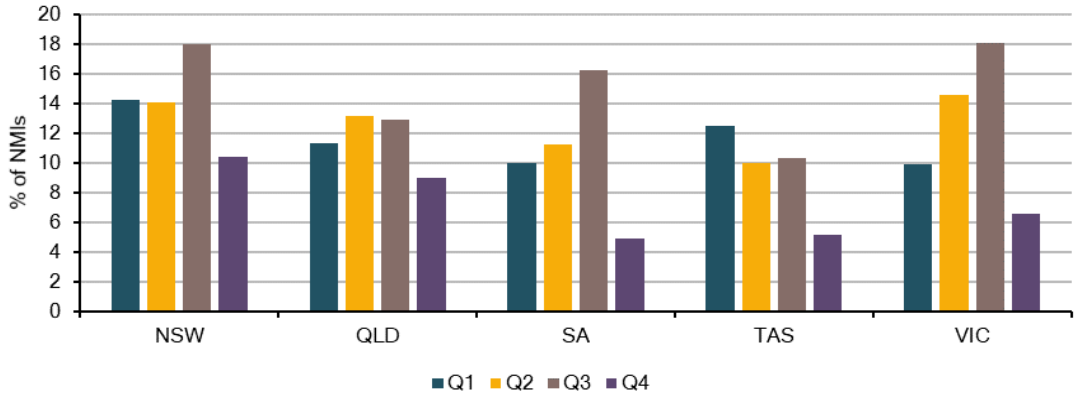
The modelling was undertaken as a means for illustrating the likely outcome of using the eligibility assessment approach discussed above to estimate the maximum potential eligibility of the NMIs within the Large Commercial and Industrial Customer segment for participation in the WDRM.

The figure below shows the maximum eligibility percentages that result in each NEM region in each quarter of 2019 from the use of the eligibility assessment methodology that was modelled.

2 This BM was found in the Phase 1 analysis to be the one that would produce the lowest median RRMSE scores if a single BM approach was to be adopted for use in the WDRM in all NEM regions, customer segments and times of day.

3 The AEMC's Rule change stated that the level of accuracy required in the WDRM should be at least as stringent as that used in the RERT. The RERT uses 0.2 RRMSE as its threshold for acceptable accuracy.

Percentage of Segment F NMIs with 90% of median RRMSEs LE 0.1 by region and quarter (CY2019)



Caveats and areas for further consideration

The impact of when the eligibility assessment is undertaken

Examination of the results in the figure above shows that the percentage of NMIs meeting the eligibility assessment criteria used in the modelling is significantly lower in Q4 that in any of the other quarters in each of the NEM regions. This could have critical implications for the program given that Q4 could be when eligibility is established for participation in the WDRM in Q1. The low percentage of eligibility in Q4 would tend to reduce eligibility in Q1, which is significant because just a very proportion of the high price events tend to occur in Q1[4].

This could be an artefact of changes in the operations of a proportion of the businesses in this segment due to the holiday season (which could result in a slowdown of activity in the last half of December preceded by a ramp up of activity in late November and early December to complete orders prior to the slowdown) and/or variability in weather from the early to late parts of the quarter).

While further analysis could be undertaken to determine why the eligible percentage of these customers is so much lower in Q4 that in Q1, AEMO could consider other options for overcoming this problem including:

- Reducing the number of days in the eligibility assessment[5], or
- Keeping the same (or similar) number of days in the eligibility assessment timeframe, but shifting the period used for eligibility for Q1 earlier in order to avoid December (or other period found to be problematic).

4 As noted in the Phase 1 analysis, Q! accounted for 50.4% of all of the high-price events that occurred in 2019.

5 While the number of days could be reduced, demand response programs in several other jurisdictions use a number of days similar to the number of weekdays in a quarter. A full quarter was used in this analysis because that was how the metering data had been organised in the Phase 1 analysis, and re-organisation of the data files was beyond the resources and timeframe available for the Phase 2 analysis.

The impact of eligibility being based on both morning and afternoon/evening price windows

Further analysis identified that requiring the candidate NMI to achieve a RRMSE of 0.1 in both the morning and afternoon/evening price windows of the day (i.e., the results shown in the figure above) resulted in maximum eligibility percentages in each quarter in every NEM region being between 4 and 5 percentage points lower than they would be if eligibility were to be based on just the afternoon/evening window.

Consideration could be given to the afternoon/evening window being used as the preferred window for assessing eligibility. This could be considered to be appropriate based on the fact that the half hours ending from 4pm to 8pm accounted for 55% of all high-price (half-hours with prices above $270/MWh) in the NEM in 2019.

Other possible approaches for eligibility assessment

The eligibility assessment approach modelled in the Phase 2 analysis was an extension of the methodology used in Phase 1 to identify the candidate BM that could be expected to be most applicable to the load profiles exhibited by NMIs in the size range that are to be eligible for participation in the WDRM.

Other approaches could be considered, each with potential advantages and disadvantages, including:

- **Use of a single RRMSE metric calculated as an average over the period** – Calculating the RRMSE across the days within the eligibility assessment period is an alternative to the daily method described above. PJM uses this sort of approach in qualifying sites that are seeking to use DR in energy trading[6].

  In our view, the key difference between the average approach used by PJM and the approach modelled in this report is the likely distribution of daily RRMSEs over a period of time. In the daily assessment, some specified proportion of the days (e.g., 100%, 90%, etc) are required to achieve an RRMSE less than a specified value. In the assessment across a period, by contrast, the site is required to meet the specified RRMSE on average[7].

  In a normal distribution the median and the average value will be the same; that is, the expected value of a BM outcome for each day would be equally likely to be above or below the average.

  However, the distribution of RRMSE values over a specified period of time is unlikely to be normal. This was indicated in the Phase 1 analysis, where a long tail of RRMSE and ARE values higher than the median was shown in the results. The use of the mean as compared to the median as the threshold in a distribution of this kind will result in more NMIs being assessed as eligible to participate.

- **Nested approaches** -- Neither of the single metric approaches described provide a cap on the error that could be experienced in the application of the selected BM in settlement of the WDRM.

  In the approach modelled, although historical data would suggest that the error in the program would be expected to be within the RRMSE criterion used on 90% of the events, the method provides no limit on the potential error the other 10% of the time. Similarly, in the

6 After the Phase 2 analysis was completed AEMO provided a copy of an internal procedure similar to this approach that they report has been used in the RERT. Limitations in project time and budget precluded modelling the results of this approach.

7 That is, the arithmetic mean of the MSE results must result in an RRMSE of less than the defined target RRMSE.

average method approximately half the events would be assumed to be characterised with an error at or below the target RRMSE, but this approach would provide no information on the likely error in the other 50% of events.

A nested approach can provide a limit on expected error. For example:

- The NMI could be required to achieve an RRMSE of 0.1 or less on 90% (or some other value) of the days in the eligibility assessment AND
- An RRMSE of 0.2 or less on ALL days within the period.

Such a formulation would bound the expected error at an RRMSE of 0.2

It should be noted that:

- All three of the methods discussed in this report are consistent with the Verification approach described in Schedules 2 and 4 of AEMO's *Reserve Contract (Long Notice)* and the RERT Panel Agreement
- Statistical accuracy is not the only issue that AEMO needs to consider in selecting a BM and specifying eligibility requirements for the WDRM. Practical considerations (which are discussed in the body of tis report) and the expectations of relevant stakeholders also need to be taken into consideration.

  This report focuses more narrowly on the implications for accuracy of different BMs and eligibility assessment approaches in order to assist AEMO in making decisions to guide the initial implementation of the WDRM. The AEMC's Rule change allows for – and AEMO expects – that the number and types of BM methodologies will expand in the future, which will allow more and more different types of customers to participate in the WDRM.

The need to include a measure of bias

The AEMC Rule change requires that an explicit criterion be specified for bias.

The Average Relative Error statistic (ARE) is the metric generally used to measure bias in BMs. A study that DNV-KEMA undertook for AEMO in 2013 suggested that a median ARE in the range of ±0.01 or 1% could be considered to be 'acceptable' and a median ARE in the range of ±0.05 or 0.5% could be considered to be 'good'[8].

Results of the analysis we conducted in Phase 1 suggested that the use of an ARE in the range of +/- 0.01 would not be expected to result in any material reduction in eligibility as compared to the use of the RERT accuracy level of RRMSE less than or equal to 0.2 on its own.

8 DNV-KEMA, *Development of Demand Response Mechanism Baseline Consumption Methodology - Phase 2 Results Final Report*, October 2013, see tables 23 and 25 on pp 2-43 and 2-44.

1. Introduction

The Baseline Methodology project was undertaken in response to the Wholesale Demand Response Mechanism (WDRM) Rule change. As discussed in section 1.2 below, the project has been conducted in two phases. This report provides the results of the second phase.

1.1. Background

The AEMC issued its Final Determination and Final Rule to implement the Wholesale Demand Response Mechanism (WDRM) on 11 June 2020. That rule establishes a new category of market participant, a Demand Response Service Provider (DRSP). DRSPs will be able to bid demand response into the NEM's wholesale electricity market in direct competition with traditional generators[9]. DRSPs will be able to engage directly with customers – without the involvement of the customer's retailer – for the provision of DR under the WDRM.

As noted in AEMO's Brief, "the WDRM design allows for a single or an aggregation of demand-responsive, controllable market load connection point(s) within a region to be identified as eligible (a qualifying load), classified, scheduled, and dispatched as a Wholesale Demand Response Unit (WDRU) by the DRSP" in response to a dispatch instruction from AEMO.

The WDRM is expected to provide greater opportunity for consumers to participate in the wholesale electricity market. This is likely to increase competition in that market and thereby put downward pressure on wholesale electricity prices.

The Rule notes that baselines are required to assess the level of demand response delivered by an individual WDRU or an aggregation of WDRUs for dispatch and settlement purposes. A baseline is an estimate of what the customer's load would have been if they had not engaged in DR. Baselines are typically developed by assessing the consumption of the WDRU during the same time periods in the recent past, with varying approaches regarding the number and types of days to be used in doing so.

Under the new rule, AEMO is required to develop one or more baseline methodologies (BMs) and related baseline settings, as well as baseline metrics to be used in the eligibility assessment and compliance testing of candidate WDRUs and the settlement of WDRUs and DRSP portfolios in DR events.

This project was undertaken to provide information for use by AEMO in determining the initial baseline(s) to be used in the WDRM as well as deciding on the approach to be used in eligibility assessment and compliance testing of WDRM participants.

1.2. Scope

The scope of the project was divided into two Phases:

- Phase 1 concerned an assessment of the efficacy of a limited set of BMs, each of which was a variant of the CAISO 10 of 10 methodology that has been used for a number of years in measuring the delivery of demand response contracted by AEMO in the Reliability and Emergency Reserve Trader (RERT) mechanism. The BMs tested were specified by AEMO and included several variants proposed by OGW. The variations concerned the type of adjustment factors that were employed.

9 In doing so, the DR that is bid will be treated essentially as a scheduled load. The scheduled load category has always been a part of the NEM but until now has only been accessible by retailers and registered customers.

The efficacy of each BM was tested in two ways:

1. By assessing its median, 10th and 90th percentile scores for accuracy, bias and variability using the same metrics that were originally applied in assessing BMs for use in the RERT[10], and
2. By assessing the proportion of the individual NMIs within each of three consumption-size segments that would meet specific accuracy thresholds with a specified frequency under each of the BMs.

Each of these assessments was undertaken using three years of actual 30-minute consumption data at the NMI level.

A separate report was prepared detailing the methodology used in the Phase 1 work and the findings and recommendations that came out of that work.

- The scope of Phase 2 of the project included:
  - The development of the methodology for calculating key accuracy/bias metrics for WDRM (to be used at eligibility assessment and compliance testing) for the specific BM chosen by AEMO as a result of the analysis from Phase 1. The recommended methodology must give consideration to:
    - What data will be required from AEMO/participants to calculate the metrics.
    - What statistical/other methods will be used to calculate the metrics for individual DRSPs under the WDRM (under eligibility assessment and during compliance testing).
    - How complex/costly the developed metrics methodology would be to implement within AEMO systems.
    - Whether the metrics process requirements represent an unreasonable barrier to entry for market participants with respect to WDRM participation.
    - The potential for participants to change their behaviour to influence the values of the baseline calculation and therefore increase the calculated value of their demand response (i.e., the potential for gaming behaviour).
  - The development of appropriate threshold values for accuracy/bias metrics for the chosen BM, including:
    - Articulating the key reasoning (backed up by data/analysis) for recommended metrics threshold values.
    - Providing a high-level assessment of the likely level of WDRM participation under proposed threshold values, i.e., likely market participant exclusions/inclusions.
    - Any other considerations regarding the metrics threshold values that AEMO needs to have regards to such as moving to 5-minute settlement etc.

    Phase 2 will devise means for adapting the use of the selected BM (or BMs in the event that different variants are found to be best suited to different customer segments or different jurisdictions) for use in the implementation of the WDRM; that is, for qualifying

10 DNV-KEMA, *Development of Demand Response Mechanism Baseline Consumption Methodology - Phase 2 Results Final Report*, October 2013,

NMIs to be registered as WDRUs, and for assessing their performance in each WDRM dispatch event.

Error! Reference source not found. contains a copy of AEMO's Brief for the project. It should be noted that the Brief specified that the work for both Phases was to be undertaken on a fixed-fee basis.

1.3. Caveats and limitations

It should be noted that while several approaches for the eligibility assessment and compliance testing were identified (and are discussed in this report), only one was modelled due to budgetary and time constraints.

1.4. Organisation of this report

The remainder of this report is organised as follows:

- Section 2 describes several methodologies that could be used for assessing the eligibility of NMIs at registration and in compliance testing
- Section 3 assesses the proportion of large commercial and industrial customer (i.e., those with annual consumption between 750 MWh and 100 GWh) that would be expected to be eligible to participate under one of the methodologies
- Section 4 describes the statistics and metrics that were used in selecting the BM to be used in the WDRM and their application in eligibility assessment and compliance testing
- Section 5 discusses some of the issues that may arise for baselining in the WDRM when 5-minute settlement is introduced.

Because not all readers of this report will have read the Phase 1 report, some material from the Phase 1 report is repeated where that information is important for understanding the work and results of Phase2.

2. Methodologies for assessing eligibility of NMIs at registration and in compliance testing

2.1. When eligibility assessment and compliance testing will be undertaken

AEMO's current thinking is that:

- Customers meeting the WDRM size threshold will be able to register for participation at any time.
- Eligibility assessment testing will be undertaken when a registration request is received. The purpose of the eligibility assessment is to test the accuracy at which the BM to be used in the WDRM predicts the load of (and therefore the demand response provided by) each NMI seeking to participate in the program. NMIs whose loads achieve the threshold level of accuracy will be eligible to participate in the program; NMIs that do not achieve the required level of accuracy will be able to re-submit a registration request and be re-tested for eligibility. again.

  The eligibility assessment test will focus on those times of day when demand response is likely to be active in the wholesale market (i.e., at times historically when high price events occur; Phase 1 found that these events were most likely to occur between the hours of 7:00 to 9:00am and 3:30 to 8:00pm[11]).
- The performance of the baseline methodology may be reviewed at AEMO's discretion as needed in order to improve baseline accuracy and/or applicability.

2.2. The approaches considered

Analysis was undertaken in Phase 1 to determine the maximum potential eligibility[12] on a quarterly basis[13] of three different customer segments (defined by annual electricity consumption volume). The three eligibility thresholds tested are shown in Table 1 below.

Table 1: Eligibility thresholds tested in Phase 1

RRMSE threshold	Proportion of days in the eligibility period for which the threshold must be met
0.1	100%
0.2	100%
0.1	90%

11 All times quoted are for the NEM standard time – Australian Eastern Standard Time (AEST).

12 Because not every customer site that could potentially participate in the WDRM will choose to do so, these figures represent maximum potential participation at each of the three eligibility thresholds.

13 In practice, eligibility assessment will be undertaken upon (or very shortly after) a registration request is received and will be conducted using the prior 50 weekdays (and possibly the 20 weekend days). To simplify the modelling task, the assessment of potential eligibility was based on the weekend days in the prior calendar quarter.

It was noted at the time this analysis was undertaken that while the first two thresholds would provide an upper bound on the error that could be expected when sites deemed eligible actually participated, the third threshold level left total error unbounded (because the error associated with the remaining 10% of the test days would not necessarily be known[14]).

Two approaches were developed that could ensure that the qualification procedure could provide an upper bound of the error associated with its use. These were as follows:

- A nested approach in which NMIs seeking to participate would be required to produce:
  - An RRMSE of 0.1 or less 90% of the time, AND
  - An RRMSE of 0.2 or less 100% of the time

  for the 3:30 to 8:00pm event window on each weekday and possibly weekend day for some period of time prior to eligibility being determined.
- A nested approach in which NMIs seeking to participate would be required to produce:
  - An RRMSE of 0.2 or less 100% of the time, AND
  - A median RRMSE of 0.1 or less, AND
  - A median ARE of +/- 0.01 or less

  for the 3:30 to 8:00pm event window on each weekday and possibly weekend day for some period of time prior to eligibility being determined.

These approaches would be applied to some time period closely preceding the time of eligibility assessment. Time periods prior to eligibility considered included the weekdays and weekends in the preceding calendar quarter, the preceding 50 weekdays and the preceding 20 weekend days, and the possibility of excluding certain periods, such as the second half of December.

It should be noted that (a) at the time these approaches were developed, the preferred BM had not been selected, and (b) these approaches could be applied to any of the candidate BMs.

2.3. The approach selected for testing

2.3.1. Approach

The approach selected in consultation with AEMO for modelling in Phase 2 was a tethered approach in which NMIs seeking to participate would be required to produce:

- An RRMSE of 0.1 or less 90% of the time
- For both the 7:00 to 9:00am event window AND the 3:30 to 8:00pm event periods on each relevant day (i.e., weekday or weekend day, excluding public holidays and days of DR activation).
- Over the course of each of the previous calendar quarters (this was carried out based on calendar year 2019).

The test was conducted using a BM with a multiplicative adjustment applied in the pre-event period only and capped at 20%.

14 However, the testing procedure could be set to return the maximum RRMSE value for the 45-day test period which would provide the maximum expected error. As it stands, this threshold states that there is only a 10% chance of the RRMSE being greater than 0.1.

2.3.2. Rationale for selecting this approach

The rationale for the selected approach was as follows:

- The tethered approach provides coverage of the vast majority of the half hours in which high prices were found to occur from 2017 through 2019
- The use of a single RRMSE metric as compared to a nested approach was chosen for its simplicity and ability to be easily explained to customers and DRSPs
- The combination of a higher level of accuracy (RRMSE of 0.1 as compared to the RRMSE threshold of 0.2 that is used in the RERT) and a requirement for that level of accuracy to be met on 90% of the days included in the eligibility assessment period was seen as providing an adequate balance between accuracy[15] and the proportion of NMIs that would be able to participate in the WDRM
- The use of the preceding calendar quarter was selected because that was how the metering data had been organised in the Phase 1 analysis; re-organisation of the data files was beyond the resources and timeframe available for the Phase 2 analysis.

15 The AEMC's Rule change stated that the level of accuracy required in the WDRM should be at least as stringent as that used in the RERT. The RERT uses 0.2 RRMSE as its threshold for acceptable accuracy.

3. Estimate of potential eligibility based on the approach modelled

The maximum potential eligibility of the Large Commercial and Industrial Customer segment was estimated as a means for illustrating the likely outcome of using the eligibility assessment approach discussed above.

3.1. Results of the modelling

Figure 1 provides the results of the modelling using the selected approach for each NEM region in each quarter in 2019. Table 2, which follows the figure, shows the specific eligibility percentages in each region in each quarter.

Figure 1: Percentage of Segment F NMIs with 90% of median RRMSEs LE 0.1 by region and quarter (CY2019)

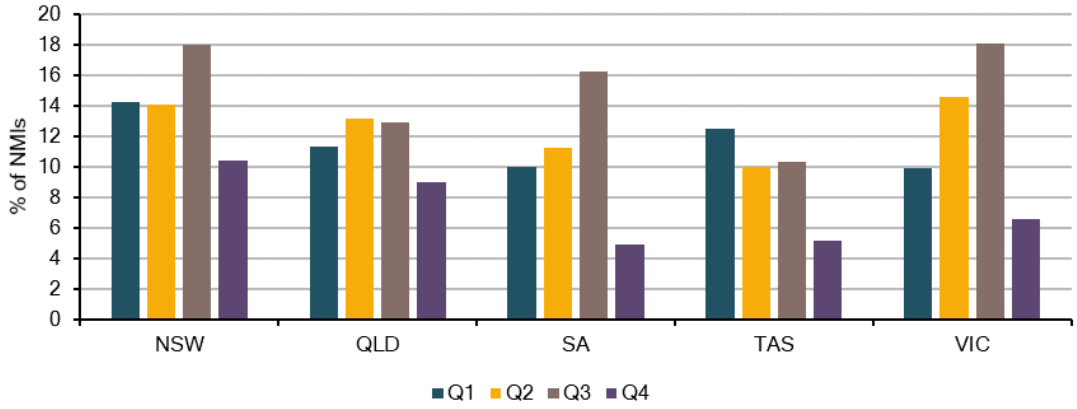


Table 2: Specific eligibility percentages by region and quarter (CY2019) shown in Figure 1

Quarter	NSW	QLD	SA	TAS	VIC	Total
1	14%	11%	10%	13%	10%	12%
2	14%	13%	11%	10%	15%	14%
3	18%	13%	16%	10%	18%	16%
4	10%	9%	5%	5%	7%	8%

It should be noted that under the proposed approach, a NMI's eligibility for the WDRM in an upcoming calendar year quarter would be determined by its outturn median RRMSE for the previous calendar quarter using the BM and acceptance criteria described above. For example, the eligibility for Q3 would be based on the results of the BM analysis for Q2.

The results above suggest one area requiring additional consideration: Eligibility percentages are significantly lower in Q4 as compared to the other quarters, which could have an impact on eligibility assessments carried out leading up to Q1. At this level of resolution[16], it is difficult to determine a cause for this difference, but it may be that the results are affected by:

16 The analysis was undertaken on a calendar quarter level. Additional analysis could be undertaken using different blocks of time, but that was not possible within the resources of this project.

- Weather variability – the spring period contains quite variable weather. This would need to be tested by a more fine-grained analysis. In addition, the second quarter, which also has variable weather, does not show this effect.
- Changes in NMI loads in the month of December – load reductions in the second half of December due to shutdowns ahead of the holiday season (possibly following load increases in late November and early December to complete orders before the holiday period) could have an impact.

The low percentage of eligible NMIs in Q4 means that the percentage of eligible NMIs in Q1 – if based on the application of baseline data for the entirety of Q4 – would likely be lower than in any other quarter. Given that the majority of the high-price events occur in Q1, this is not ideal.

AEMO should consider undertaking further analysis of:

- The specific number of days to be used in the eligibility assessment, and
- The impact of different timeframes and durations for eligibility assessments undertaken in Q4, including avoiding the pre-holiday period.

3.2. Implications of the tethered approach

Figure 2 through Figure 5 on the following pages show the percentage of NMIs that meet the RRMSE criterion in the morning and afternoon/evening windows individually and on a combined basis, by region and quarter in 2019[17].

Logically, the percentage that will meet the criterion for the combined windows cannot be higher than the percentage characterising the window that achieves the lower eligibility. This is clearly reflected in the figures that follow. More specifically, the figures show that the combined window approach will result in eligibility percentages about 5 percentage points lower than they would be if eligibility were to be based on just the afternoon/evening window.

This outcome is not surprising given the fact that the Phase 1 results showed that different BMs produced better results in the morning window as compared to the afternoon/evening window. The Additive adjustment with a 20% adjustment cap was shown to provide lower median RRMSEs in the morning window while the Multiplicative adjustment with a 20% adjustment cap was shown to provide lower median RRMSEs in the afternoon/evening window[18].

As most of the high priced events occur during the evening window, it would seem most appropriate to use the BM that produces the best accuracy (as measured by lowest median RRMSE value) for assessing eligibility.

17 Tables of these results are provided in Appendix B.

18 The Phase 1 analysis actually indicated that the Multiplicative adjustment with a 40% adjustment cap provided very similar levels of accuracy in the afternoon/evening window.

Figure 2: Specific eligibility percentages by region for quarter 1 2019

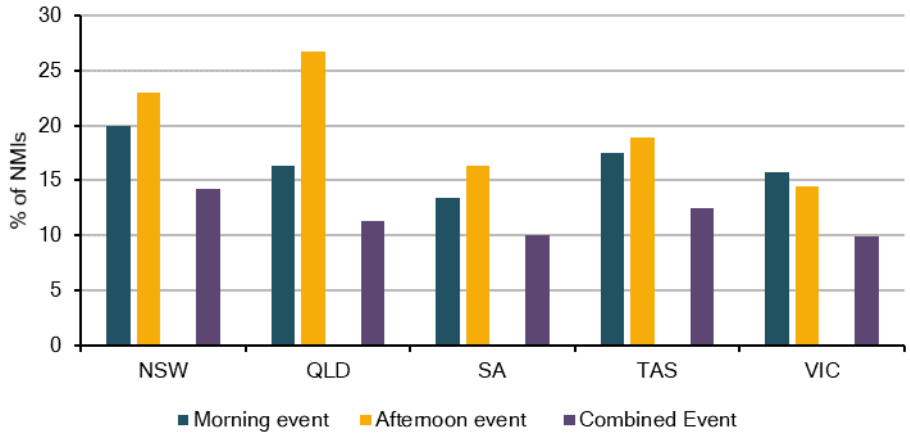


Figure 3: Specific eligibility percentages by region for quarter 2 2019

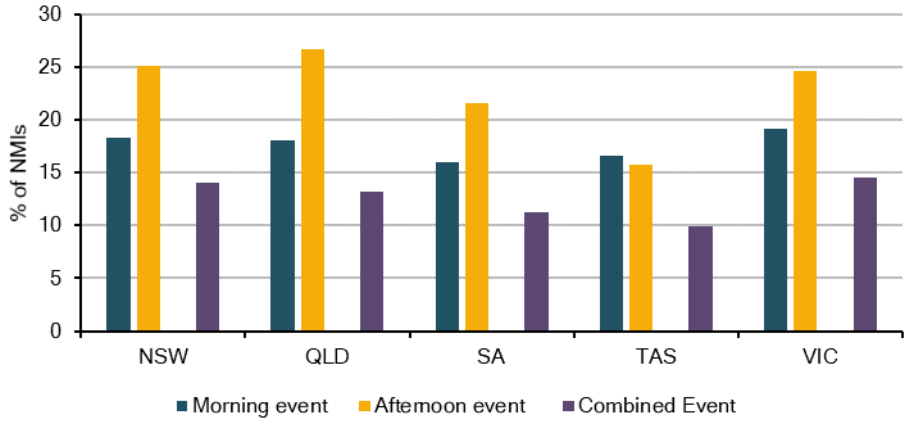


Figure 4: Specific eligibility percentages by region for quarter 3 2019

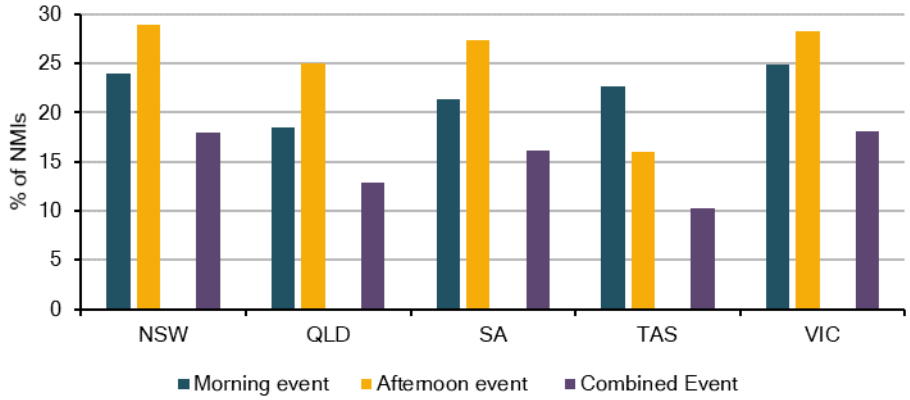
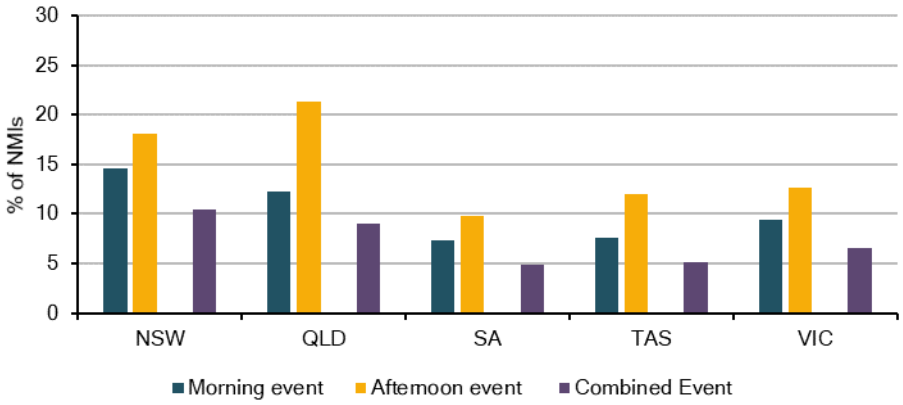


Figure 5: Specific eligibility percentages by region for quarter 4 2019



3.3. Practical considerations

AEMO's Brief asked us to comment on the following practical considerations associated with the BM methodology and its use in eligibility assessment and compliance testing:

- Data requirements – The data requirements of the proposed BM methodology and its use for eligibility assessment and compliance testing are minimal, being limited to:
  - Consumption data for the NMI in question for:
    - The period ultimately decided upon as being required for eligibility assessment
    - Construction of the baseline for settlement on event days
    - The period ultimately decided upon as being required for compliance testing.

    We note that the consumption data is readily available to AEMO and easily available to the prospective WDRU or its DRSP.
  - Evidence that the NMI meets the consumption threshold for participation in the WDRM – This is a matter of annual consumption, which is readily available to all three of the relevant parties. A decision will need to be made as to what time period this requirement applies (i.e., most recent 12 months, most recent calendar year, most recent financial year) and whether partial or limited data can be extrapolated or estimated for new sites.
- Complexity/cost with regard to integration with AEMO's systems – Although we are not familiar with the details of AEMO's systems, given the data requirements and the relative simplicity of the required processing, it is our view that integration of the BM methodology and its use in eligibility assessment and compliance testing and settlement of the dispatch of WDRUs will not pose any undue complexity or cost to AEMO. The most complex items in the processing of the methodology for any of these purposes is likely to be the de-selection of non-relevant days for the calculation being undertaken (i.e., weekdays or weekend days, public holidays, and days on which DR was activated by the site).
- Potential barriers from the perspective of potential WDRM participants – The simplicity of the 10 of 10 baseline methodology and its straightforward data requirements should make it fairly easy for WDRUs and DRSPs to understand and use in considering the potential eligibility of their sites for participation in the WDRM. The methodology is however, a fairly basic variant of the CAISO 10 of 10 methodology, which is best suited to NMIs whose load shapes are relatively consistent:
  - from day to day in the hours used for eligibility assessment and compliance testing, and

- in the intervals in which DR is actually dispatched for the NMI under the program.

This will not suit a material proportion of the customers who are eligible to participate in the WDRM by virtue of their annual consumption but it is consistent with the requirements of the AEMC's Rule.

Further discussion of this aspect of the methodology is provided in section 3.4 below. It is important to note that the Rule change allows for – and AEMO expects – that the number and types of BM methodologies will expand in the future, which will allow more and more different types of customers to participate in the WDRM.

3.4. Inclusions and exclusions when using the simple approach.

As noted above,

- Simple 10 of 10 BM methodologies are most readily applicable to end-use facilities whose load shapes are relatively consistent from day to day during the hours used for eligibility assessment and compliance testing, and in which DR is actually dispatched by the facility under the program
- This will not suit a material proportion of the customers who are eligible to participate in the WDRM by virtue of their annual consumption, but whose load shapes are more variable.

Examples of the types of customers that will be excluded from participation by the BM to be used include those:

- Whose loads exhibit significant sensitivity to temperature
- That have a DER system (and particularly a PV array) whose generation capacity is a material proportion of the facility's average demand
- Whose operating schedules vary from day to day, even if that variation is consistent from week to week (e.g., a facility that operates certain loads only on specific days of the week)
- With intermittent and irregular internal loads that are a material proportion of the facility's average demand
- That routinely operate their plant in a price-responsive manner.

4. Statistical approaches and metrics

This section of the report is an edited excerpt from the Phase 1 Report. It discusses the specific metrics that will be used in setting the performance thresholds to be used in the program for settlement, eligibility assessment and compliance testing.

4.1. Metrics used during the analysis

This study used the same metrics that were used in the study undertaken for AEMO by KEMA in 2013 to assess the accuracy, bias and variability of the candidate BMs. Briefly, these are:

- Relative Root Mean Square Error (RRMSE), which is a measure of the accuracy of the baseline. It is a measure of the differences between the half-hourly consumption predicted for a NMI and the consumption actually observed.

  Accuracy refers to how well the BM represents the true counter-factual for the event period; that is. how well it removes the "noise" of daily variations. It is a combination of precision and bias (as discussed further below and illustrated in Figure 6). The RRMSE statistic measures both precision and bias, so a low RRMSE result is likely to be correlated with low scores on both the ARE and RER[19].

  Based on the threshold specified in AEMO's Reserve Contract (Long Notice), AEMO's RERT Panel Agreement and the AEMC Rule change, an RRMSE of 0.2 or lower has been used as the threshold of 'acceptable' accuracy, and an RRMSE of 0.1 or lower as the threshold of 'good' accuracy.

- Average Relative Error (ARE) is a measure of bias derived by adding the difference between the half-hourly consumption figures in the baseline and the actual load for each day. The closer the ARE is to zero, the closer the baseline is to being unbiased.

  A positive bias indicates that the baseline will tend to over-estimate the amount of demand response delivered, while a negative ARE indicates that demand response will be under-estimated. For example, a median ARE value of +0.01 would mean that the BM method would overestimate the DR provided by 1%.

  The RERT does not currently use the ARE metric as a specific acceptance criterion, but an assessment of bias is required by the AEMC Rule change.

  In Phase 1 we used our interpretation of the results of the study DNV-KEMA undertook for AEMO in 2013 – that a median ARE in the range of ±0.01 or 1% could be considered to be 'acceptable' and a median ARE in the range of ±0.05 or 0.5% could be considered to be 'good'[20] – in assessing the various candidate BM methodologies for use in the WDRM. These same criteria would seem applicable for use in the WDRM.

- Relative Error Ratio (RER), which is a is a measure of the precision of the baseline. It is derived by comparing the standard deviation of the baseline's prediction errors as fraction of the average load. The smaller the median RER, the less variable a baseline's error is for the typical customer and therefore the better the baseline performs across a wide variety of circumstances.

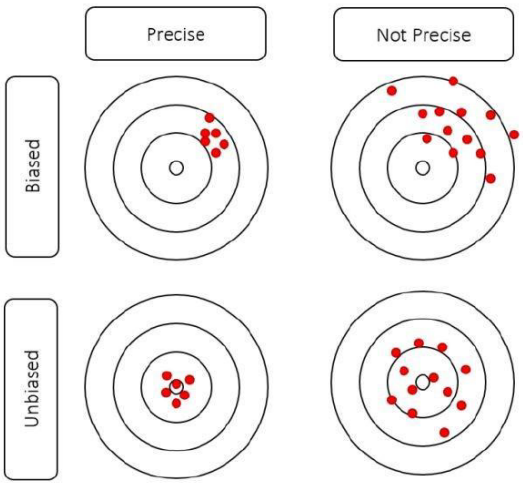
19 This can be seen in the Phase 1 results; see particularly the tables in sections 3, 4 and 5.

20 DNV-KEMA, *op. cit.*, see tables 23 and 25 on pp 2-43 and 2-44.

Neither AEMO, in contracting the use of load reduction for reserve purposes, nor the AEMC Rule change, uses the RER as an eligibility criterion. For this reason and because the information provided by the RER is also provided by the RRMSE calculation, the RER was calculated in the Phase 1 analysis, but it will not be used in assessing NMI performance in settlement, eligibility assessment or compliance testing.

The four 'targets' in Figure 6 below provide a graphic representation of the relationship of bias and precision (variability) to accuracy (the bullseye within the target):

Figure 6: BM metrics measure the accuracy, bias and precision of candidate baselines



Further information on the mathematical formulations and their use is provided in the Functional Specification presented in Appendix B of the Phase 1 Report.

4.2. Potential approaches to applying the statistics and metrics to eligibility assessment and compliance testing in the WDRM

Eligibility assessment/compliance testing is used to accept a customer or site into the WDRM. It is a different activity than the settlement of a specific event.

The fundamental difference between the two is that the BM is used in settlement as a yardstick to determine the amount of DR delivered. By contrast, in eligibility assessment or compliance testing the accuracy produced by the use of the BM is used to determine whether the BM can predict the load at that NMI with sufficient accuracy for the purposes of the WDRM itself.

Once the NMI's load has been shown to be able to be predicted at the selected level of accuracy in eligibility assessment, there is confidence that its application in settlement will be characterised by a similar level of accuracy. The eligibility assessment approach can be applied at any point after the NMI has been accepted in the program to re-check that the NMI's load continues to be able to be predicted with suitable accuracy using the selected BM.

The sections below discuss two approaches for how the metrics within the BM can be used in eligibility assessment/compliance testing. Both are consistent with the Verification approach described in Schedules 2 and 4 of AEMO's *Reserve Contract (Long Notice)*, and in the RERT Panel Agreement. The discussion of each of these approaches focuses on the RRMSE statistic. As noted earlier, the AEMC Rule change requires that an explicit criterion will also need to be specified for bias (the ARE). The Phase 1 analysis suggests that the use of an ARE range of +/- 0.1 (which has generally been seen as being acceptable in other studies) would not be expected to result in any material reduction in eligibility as compared to the use of the RERT accuracy level of RRMSE less than or equal to 0.2 on its own.

4.2.1. Application of the metrics on a daily basis for a specified period

In assessing the maximum potential eligible proportion that would be provided by the application of the RRMSE metric on a daily basis for a specified number of days (as presented in section 3.1, we used the same approach for assessing the predictive accuracy of the NMI's load as we used in assessing the accuracy of the candidate BMs in Phase 1. That is, we predicted each NMI's load using the selected BM for each day of the specified period (in this case a calendar quarter). The RRMSE for both the morning and afternoon/evening event periods on each day was calculated and compared to the threshold RRMSE value (in this case 0.1). Where the RRMSE was less than or equal to 0.1 the day was counted towards the eligibility assessment requirement. Where 90% of the relevant days within the calendar quarter achieved the target RRMSE, the NMI was deemed eligible to participate in the WDRM.

The proportion of customers that could potentially be eligible to participate in the WDRM can be fine-tuned by changes in the threshold value used for either (or both) the RRMSE or the frequency at which that value needs to be attained on a daily basis.

4.2.2. Application of the metrics across the days within a period

Calculating the RRMSE across the specified period is an alternative to the daily method described above. PJM uses this sort of approach in qualifying sites that are seeking to use DR in energy trading. The PJM approach works as follows:

- Identify the relevant 45 days to be tested – defined as non-event days that were not holidays or, in the case of a weekday, weekend days
- Calculate the baseline using the 10 of 10 BM for each of the days and develop the Mean Square Error for each of the days
- Sum the Mean Squared Errors over the 45 days and divide by 45, which results in the average of the Mean Squared Errors
- Take the square root of the average Mean Squared Error and divide it by the sum of the demands for the 45 days to develop a Relative Root Mean Squared Error for the 45-day period.

This results in a single RRMSE for the 45-day period which can be compared to the threshold RRMSE to determine the site's eligibility to use DR in energy trading. Fine-tuning of the proportion of customers that could potentially be eligible to participate can only be undertaken by changing the threshold value for the average RRMSE score.

4.2.3. Commentary and comparison of the approaches

Comparison of the two approaches

Conceptually, the daily approach is more like the approach we used in Phase 1 for assessing the likely level of accuracy that each of the candidate BMs would produce if applied to actual historical load data; that is, it applies the BM to the load on specific days. As such, it provides a measure of the BM's ability to predict the NMI's consumption on a day-to-day basis over the specified period, and therefore can be used to provide a distribution of the accuracy of those daily predictions. For simplicity, we used the median RRMSE in Phase 1 as a single measure of that distribution but also examined the range and percentile accuracy of the daily predictions.

By contrast, the primary output metric of the PJM approach is a single RRMSE calculated as the average error over the period[21].

In our view, the key difference in these two approaches is the likely distribution of daily RRMSEs over a period of time. In the daily assessment, some specified proportion of the days (e.g., 100%, 90%, etc) are required to achieve an RRMSE less than a specified value. In the assessment across a period, by contrast the site is required to meet the specified RRMSE on average[22].

In a normal distribution the median and the average value will be the same; that is, the expected value of a BM outcome for each day would be equally likely to be above or below the average.

However, the distribution of RRMSE values over a specified period of time is unlikely to be normal. This was indicated in the Phase 1 analysis, in which it was clear that the distribution of daily RRMSE values is materially skewed to the right (that is, they tend to have a long tail of relatively high RRMSE values). In such a distribution, the mean of the daily values will be higher than the median. Therefore, the use of the mean as compared to the median as the threshold will result in more NMIs being assessed as eligible to participate[23].

In our view the daily assessment approach is preferable to the average assessment approach because it directly assesses the frequency at which the load of the NMI will be predicted within the chosen accuracy bound by the BM. The average over the period approach cannot do this. To be more specific, the eligibility assessment criterion tested in this report – that the NMI produces an RRMSE of 0.1 against the BM for 90% of the relevant days within the specified period – means that there is only a 10% chance that the RRMSE will exceed 0.1[24].

21 It should be noted that, because the metric is developed from the daily Mean Squared Error, the distribution of that statistic, or a daily RRMSE, could be examined as part of the registration process (though there is no indication that this is done by PJM).

22 That is, the arithmetic mean of the MSE results must result in an RRMSE of less that the defined target RRMSE.

23 The corollary to this is that for any desired proportion of eligible customers, the average RRMSE value would need to be set lower than the corresponding value for the median RRMSE.

24 A similar approach could be used as an adjunct to the average RRMSE based on the distribution of the NMI's daily Mean Squared Error scores.

Comparison with the approach used in the RERT

It is important to note that the approach being considered for eligibility assessment testing in the WDRM differs from that used in the RERT in two important ways:

1. In the RERT, DR that is to be provided by more than one end-use customer can be registered, as a portfolio[25]. Registration is undertaken through a Tender process in which the proponent (whether a DR aggregator or a single site) provides detailed information on the nature of the DR it can provide, including the specific NMIs included in the portfolio and the nature of the DR to be provided by each. Over-recruitment within the portfolio is a common strategy for providing confidence to AEMO and the proponent that the amount of DR being offered for contract will be delivered when called.
2. Proponents in the RERT are required to demonstrate through a test activation that they can deliver the amount of DR for which they have been contracted. If the contracted amount is not provided in the test activation, the proponent can request a re-test. Where neither test meets the contracted amount AEMO can reduce the contracted amount to the higher of the two test results.

It is not clear at this point what if any information DRSPs will be required to provide about the nature of the DR and other characteristics and the NMIs within its aggregation upon application for eligibility assessment for the WDRM, but it is known that each NMI will be required to demonstrate its ability to produce RRMSE and ARE scores that meet specific levels over a specified number of days. (This is clearly a more stringent test than the result of the single-day, portfolio-based test used in the RERT.)

4.2.4. Considerations in choosing an approach

The differences discussed above between how eligibility assessment is done in the RERT and how the Rule change requires it to be done in the WDRM mean that a direct transfer of the accuracy requirement of the RERT to the WDRM will actually result in the WDRM imposing a materially higher level of accuracy for participation by a NMI than that required in the RERT. This should be taken into account in the final setting of the WDRM eligibility assessment requirement.

Another important consideration has to do with the importance of a Type 1 versus a Type 2 error in how the eligibility assessment process is set[26]. In our view, avoidance of a Type 2 error is the more important consideration from the perspective of the market. This is because once allowed to participate in the program, the baseline will be used measure and determine payment for the DR provided by the NMI. Where the eligibility assessment process allows NMIs for which the BM is less accurate to participate, that reduced accuracy will be accepted in settlement.

25 DR aggregations in the RERT are also settles as a portfolio and are compliance tested (if needed) as a portfolio. These options would not be available to DR offered by a single site that meets the RERT's 10MW minimum DR contribution requirement.

26 In this context, a Type 1 error is rejecting a NMI whose load can be accurately predicted and should therefore be included in the WDRM and a Type 2 error is accepting a NMI whose load cannot be accurately predicted by the BM method and should therefore be excluded. The Type 2 error will increase the risk of error in dispatch and settlement, noting that that error could constitute either an over- or under-payment to the DR provider.

Overall, it is our view that a eligibility assessment approach that uses a threshold value for the RRMSE and a specified proportion of days on which that value is to be met during the relevant days in the specified period[27] is the superior approach from an accuracy perspective. Such an approach will be better at avoiding Type 2 errors than the average RRMSE approach simply because it includes a measure of the expected frequency at which the target RRMSE will not be met. That is not considered in the approach that uses the average RRMSE over the specified period as that approach is generally applied.

We do note, however, that the single threshold approach discussed in Section 3 does not bound the total error possible due to its use. As noted in Section 2.2, there are approaches that can provide such a bound. They were not tested in this Phase 2 work due to limitations of time and budgetary under which preference was given to a simpler approach. In our view, it would be worth further consideration of the degree of stakeholders' perceptions of the added complexity of the alternative approaches discussed in Section 2.2 as compared to the advantages it provides in accuracy. In this regard we note that this is really a matter of perception as there is very little difference in the complexity of the calculations that would be needed for the two eligibility assessment approaches.

Finally, we also note that statistical accuracy is not the only issue that AEMO needs to consider in selecting a BM and specifying eligibility requirements for the WDRM. This report was commissioned to assess the implications for accuracy in different BMs and eligibility assessment approaches for use by AEMO in making decisions to guide the initial implementation of the WDRM. As noted earlier, the Rule change allows for – and AEMO expects – that the number and types of BM methodologies will expand in the future, which will allow more and more different types of customers to participate in the WDRM.

27 Including a nested approach in which the RERT accuracy level of 0.2 RRMSE is required to be met 100% of the time.

5. Potential issues arising from 5-minute settlement

It is difficult to be definitive about the issues that could potentially arise from the shift to 5-minute settlement. Assuming that the BM would then also move to the use of 5-minute intervals, it is clear that:

- The calculation processes for eligibility assessment, settlement and compliance testing will need to be modified, though this should not be particularly difficult.
- There will need to be a transition period in which although metering and settlement will have already moved to 5-minute basis, eligibility assessment and to for a shorter period the construction of baseline will need to continue to be undertaken on the 30-minute basis.
- The amount of data needing to be processed will increase proportionally, though it is unlikely that this would pose a material barrier.
- It is possible that loads that vary significantly within a half hour will exhibit materially higher RRMSEs under the 5-minute calculation than the current 30-minute calculation. This is illustrated in the four cases shown in Table 3 below. In each case the 30-minute error for this hypothetical NMI is the same (i.e., the sum of the errors in the first line of each case sums to 30). But, as can be seen, the Root MSE (RMSE) value differs materially with the distribution of those errors over the six 5-minute intervals within the half hour.

Table 3: Impact of different distributions of error by 5-minute interval on RRMSE

Case		5-minute interval						Mean Square Error (MSE)	Root MSE
		1	2	3	4	5	6		
1	Error	5	5	5	5	5	5		
	Squared Error	25	25	25	25	25	25	25	5.0
2	Error	10	4	4	4	4	4		
	Squared Error	100	16	16	16	16	16	30	5.5
3	Error	20	2	2	2	2	2		
	Squared Error	400	4	4	4	4	4	70	8.4
4	Error	25	1	1	1	1	1		
	Squared Error	625	1	1	1	1	1	105	10.2

Without real data it is not possible to estimate how many NMIs are likely to exhibit this effect or to what extent, and therefore the degree to which it could change the proportion of customers eligible to participate in the WDRM at any target RRMSE. However, it is reasonable to assume that the shift to 5-minute settlement (and the accompanying shift to BMs using 5-minute intervals) will have the effect of increasing RRMSEs for some NMIs, leave some NMIs relatively unchanged, but improve the RRMSEs of no NMIs. This could be addressed by either adjusting the relevant RRMSE value or by reducing the proportion of days within the specified period that a site must achieve the desired RRMSE.

Appendix A: AEMO Project Brief

Schedule 1 – Consultancy Services

1. Description of Consultancy Services, Deliverables

AEMO wishes to engage a consultant with skills and knowledge relating to electricity markets and demand response methodologies, to test and give advice on a variety of baseline methodologies under the proposed Wholesale Demand Response Mechanism (WDRM).

Background

On 12 March 2020, the AEMC released a second draft determination and draft rule to implement a WDRM.[1] Under the second draft rule, a new category of registered participant, a demand response service provider (DRSP), would be able to bid demand response directly into the wholesale market as a substitute for generation. A DRSP could also engage directly with a customer without the involvement of that customer's retailer.

The mechanism is designed to provide greater opportunities for consumers to participate in the wholesale market by bidding in demand reductions as a substitute for generation, thereby unlocking under-utilised demand response in the national electricity market (NEM). The WDRM design allows for a single or an aggregation of demand-responsive, controllable market load connection point(s) within a region to be identified as eligible (a qualifying load), classified, scheduled, and dispatched as a Wholesale Demand Response Unit (WDRU) by DRSP.

In June 2020, the AEMC approved a final rule determination relating to the wholesale demand response mechanism.

Baselines

The demand response settlement process requires the establishment of a baseline for each single WDRU. Baselines are an estimate of the consumption per trading interval during a day, based on a history of like days in the near past. Baselines are required in the draft Rule for two main purposes:

- They are the counterfactual energy amount for each single WDRU that is dispatched individually or as part of an aggregated WDRU for demand response. This baseline is required for demand response settlement.
- They are the counterfactual energy amount for the WDRU that is dispatched for demand response.

Under the proposed Rules, AEMO must develop one or more baseline methodologies (BLMs) and related baseline settings, as well as baseline methodology metrics for eligibility/compliance testing.

DNV KEMA analysis

In 2013, AEMO engaged consultants DNV KEMA to provide advice on the construct and concepts surrounding the development and implementation of a baseline consumption methodology for the implementation of the Demand Response Mechanism in the NEM. The study was conducted in two phases. Phase 1 includes research into the baseline methodologies in use at the various United States (US) Independent System Operators (I15 SOs). Phase 2 involves the testing of the efficacy of potential baseline consumption methodologies for use in NEM. This work was used as the basis for selecting the

[1] AEMC Wholesale Demand Response Mechanism – Rule Change: https://www.aemc.gov.au/rule-changes/wholesale-demand-response-mechanism

methodology (CAISO "10 of 10" with additive adjustment) currently used for determining baselines under the Reliability and Emergency Reserve Trader (RERT) mechanism[2].

The CAISO "10 of 10" methodology is AEMO's starting point for developing a BLM for the WDRM. The analysis in Phase 2 of the report is the basis for developing a large part of the scope of work (specifically around the accuracy/bias and variability metrics). The DNV KEMA reports can be accessed at:

- Phase 1 Report - https://www.aemo.com.au/Electricity/National-Electricity-Market-NEM/Security-and-reliability/-/media/FE9ABE8C64064E1E903154D3C18ADFA4.ashx
- Phase 2 Report - https://www.aemo.com.au/Electricity/National-Electricity-Market-NEM/Security-and-reliability/-/media/E146E007B78447C685201A3A42F1EC8D.ashx

Scope of work

AEMO envisages this work to be undertaken over 2 phases.

- Phase 1 - Baseline Methodology Testing
- Phase 2 - Baseline Methodology Metrics

Phase 1 - Baseline Methodology Testing

AEMO is seeking a consultant to test the efficacy a variety of "RERT like" BLMs with commonly accepted adjustment approaches. The analysis is to use multiple metrics and recent NEM data for a range of potential WDRM participants. BLMs to be analysed are a restricted to the following (for both weekday and weekend scenarios):

i. BLM – CAISO 10 of 10 - no adjustment
ii. BLM - CAISO 10 of 10 - additive adjustment no cap
iii. BLM - CAISO 10 of 10 - additive adjustment with % cap
iv. BLM - CAISO 10 of 10 - multiplicative adjustment
v. BLM - CAISO 10 of 10 - multiplicative adjustment with cap

Phase 2 - Baseline Methodology Metrics

Phase 2 of the analysis is to involve:

- The development of the methodology for calculating key accuracy/bias metrics for WDRM (to be used at registration and compliance testing) for the specific BLM chosen by AEMO as a result of the analysis from Phase 1.
- The development of appropriate threshold values for accuracy/bias metrics for the BLM.

Key Deliverables – Phase 1 - Baseline Methodology Testing

1) Recommendation to, and agreement with, AEMO around key analysis parameter required for producing outputs that AEMO can use to make key decisions around the suitability of a BLM for WDRM:
    - Number of candidate days used.

[2] For a full description of the baseline methodology employed under the RERT scheme, refer to Appendix F in the following document: https://arena.gov.au/assets/2019/06/demand-response-funding-announcement-update-2.pdf

- Candidate day conditions to be considered (i.e. regional load characteristics, weather, pricing etc.).
- Trading intervals included.
- NMI segmentation (size, industry etc)
- Number of NMIs required for statistically significant results (sample size of data).
- Data sampling considerations (random, targeted etc)
- Statistical method(s) for analysis of metrics and useful outputs (point/distribution statistics).
- Any other relevant parameters.

2) Use metrics to establish the BLM's statistical properties by agreed NMI segmentation. The metrics to be used are:
    - Accuracy - how closely a baseline methodology predicts customers' actual loads in the sample.
    - Bias - the systematic tendency of a baseline methodology to over- or under-predict actual loads.
    - Variability - how well the baseline methodology predicts half hourly load under many different conditions and across many different customers. As variability is not part of WDRM metrics, this metric may represent a small part of the analysis, depending on data availability.
3) Rank of BLMs relative to each other, together with a high-level advice/recommendation as to which BLM would be most suited for the implementation of WDRM. For each BLM, consider:
    - Relevant baseline settings.
    - Gaming opportunities.
    - Feasibility of implementing/administering methodology under the WDRM Rules.
    - Suitability of each BLM to each market segments (i.e. who would likely be in or out).
    - Likely cost/benefit implications (AEMO/participants).
4) High level assessment of WDRM participation under each BLM, considering:
    - Likely baseline metrics (determining markets segments eligibility).
    - Any participant barriers to entry.
5) The above key deliverables (analysis and data) to be given to AEMO in the form of a professional consultant's report.
6) AEMO is to be supplied with a soft copy of the raw outputs of the analysis for each NMI and each BLM and each tested TI (i.e. the underlying data used to compute the summary statistics by segmentation).

Key Deliverables - Phase 2 - Baseline Methodology Metrics

1) The development of the methodology for calculating key accuracy/bias metrics for WDRM (to be used at registration and compliance testing) for the specific BLM chosen by AEMO as a result of the analysis from Phase 1. The recommended methodology must give consideration to:
    - What data will be required from AEMO/participants to calculate the metrics.
    - What statically/other methods will be used to calculate the metrics for individual DRSPs under the WDRM (under registration and during compliance testing).

- How complex/costly the developed metrics methodology would be to implement within AEMO systems.
- Whether the metrics process requirements represent an unreasonable barrier to entry for market participants with respect to WDRM participation.

2) The development of appropriate threshold values for accuracy/bias metrics for the chosen BLM, including:
    - Articulating the key reasoning (backed up by data/analysis) for recommended metrics threshold values.
    - Providing a high-level assessment of the likely level of WDRM participation under proposed threshold values, i.e. likely market participant exclusions/inclusions.
    - Any other considerations regarding the metrics threshold values that AEMO needs to have regards to such as moving to 5-minute settlement etc.
3) The above key deliverables (including any analysis and data) to be given to AEMO in the form of a professional consultant's report.

Data provided by AEMO

AEMO is to provide the following data to consultant at start of project:

- Historical 30-minute meter data (exact data sample, dates and breakdown to be discussed with consultant).

Other data that may be provided by AEMO depending on discussion with consultant:

- Historical regional load data.
- Historical regional price data.
- Historical regional weather data.
- Data on RERT activation, Demand Side Participation, Directions and FCAS activation.
- NEM Demand side participation (DSM) Portal data.
- Historical Scada 5-minute load data.

2. Deliverables Summary

The Consultancy Services includes the provision of the following Deliverables:

Deliverable	**Description**
Phase 1	
Key analysis parameters	Recommendation to, and agreement with, AEMO around key analysis parameter required for producing outputs that AEMO can use to make key decisions around the suitability of a BLM for WDRM.
BLM statistical properties	Use metrics (accuracy and bias) to establish the baselines' statistical properties by agreed NMI segmentation.

Deliverable	Description
BLM rankings	Rank of BLMs relative to each other, together with a high-level advice/recommendation as to which BLM would be most suited for the implementation of WDRM.
WDR participation assessment	High level assessment of WDRM participation under each BLM.
Final Consultant's Report	Analysis and data to be given to AEMO in the form of a professional consultant's report.
Analysis data	AEMO is to be supplied with a soft copy of the raw outputs of the analysis for each NMI and each BLM and each tested TI.
Phase 2	
Process for calculating metrics	The development of the methodology for calculating key accuracy/bias metrics for WDRM for the specific BLM chosen by AEMO as a result of the analysis from Phase 1.
Metrics threshold values	The development of appropriate threshold values for accuracy/bias metrics for the chosen BLM.
Final Consultant's Report	Analysis and data to be given to AEMO in the form of a professional consultant's report.

Appendix B: Eligibility percentages, Large Commercial and Industrial Customer Segment, by region and quarter 2019

Table 4: Specific eligibility percentages, Large Commercial & Industrial Customers, by region, Q1 2019

	NSW	QLD	SA	TAS	VIC
Morning event	20.0	16.3	13.4	17.5	15.8
Afternoon event	23.0	26.7	16.3	18.9	14.5
Combined Event	14.2	11.4	10.0	12.5	9.9

Table 5: Specific eligibility percentages, Large Commercial & Industrial Customers, by region, Q2 2019

	NSW	QLD	SA	TAS	VIC
Morning event	18.3	18.1	16.0	16.7	19.2
Afternoon event	25.1	26.7	21.6	15.7	24.7
Combined Event	14.1	13.2	11.3	10.0	14.6

Table 6: Specific eligibility percentages, Large Commercial & Industrial Customers, by region, Q3 2019

	NSW	QLD	SA	TAS	VIC
Morning event	23.9	18.5	21.4	22.7	24.9
Afternoon event	28.9	25.0	27.4	16.0	28.3
Combined Event	18.0	12.9	16.2	10.3	18.1

Table 7: Specific eligibility percentages, Large Commercial & Industrial Customers, by region, Q4 2019

	NSW	QLD	SA	TAS	VIC
Morning event	14.6	12.2	7.4	7.6	9.3
Afternoon event	18.0	21.4	9.7	12.0	12.6
Combined Event	10.4	9.0	4.9	5.2	6.6