Est. 2008

Oakley Greenwood

Phase 1 WDRM Baseline Methodology Analysis Results and Recommendations

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 Cruikshank (acruikshank@oakleygreenwood.com.au) Mark Darlington (mark.darlington@optimate.net.au) Dennis Lilford (dennis.lilford@optimate.net.au)
Date	December 2020

CONTENTS

FIGURES

TABLES

Executive summary

Project background and purpose

The Australian Energy Market Commission's (AEMC) Final Determination and Final Rule to implement the Wholesale Demand Response Mechanism (WDRM), published in June 2020, established 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.

DRSPs will be able to engage directly with customers and without the involvement of the customer's retailer for the provision of demand response (DR) under the WDRM. Like a generator, the DRSP will be required to enter bids specifying the quantity of DR they are prepared to offer and the price at which they are prepared to offer it for. This raises the issue of measurement, as DR, unlike energy generation, cannot be measured directly. Rather, it must be estimated by comparing actual consumption with a prediction of what would have occurred if the request for DR had not been made. There are several approaches for generating this prediction, of which baselining- using the history of the site's demand – is the most common.

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 customer's consumption 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 registration and compliance testing of customer sites that want to provide DR, and the settlement of those customers and DRSP portfolios in when the DR they bid into the wholesale market is dispatched.

The project scope comprised two phases:

- The objective of Phase 1 was to identify which of a limited set of BMs could be expected to provide the greatest level of accuracy when assessing for settlement the DR delivered by customers participating the WDRM. The BMs considered were specified by AEMO and were all variants of the CAISO 10 of 10 methodology that has been used for a number of years in measuring the delivery of DR contracted by AEMO in the Reliability and Emergency Reserve Trader (RERT) mechanism.
- 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 NMIs to be registered as Wholesale Demand Response Units (WDRUs), and for assessing their performance in each WDRM dispatch event.

Phase 1 approach

Baseline methodologies considered

AEMO's Brief specified testing of the following BMs[1]:

[1] 'No adjustment' and 'no cap' options were also specified in the Brief but were not pursued as they were considered to be of low value compared to the other options. Options with a post-period as well as a pre-period adjustment were also assessed but were dropped from consideration for implementation due to the opportunity they pose for gaming.

- CAISO 10 of 10 - pre-period additive adjustment with 20% cap
- CAISO 10 of 10 - pre-period additive adjustment with 40% cap
- CAISO 10 of 10 - pre-period multiplicative adjustment with 20% cap
- CAISO 10 of 10 - pre-period multiplicative adjustment with 40% cap

The use of a pre-period adjustment refers to a methodology for adjusting the consumption level of the baseline to reflect the consumption level of the customer site in the time period preceding DR activation on the price event day. The additive approach makes this adjustment in absolute terms (that is, the difference in energy consumption levels is added or subtracted from the original baseline). The adjustment in the multiplicative approach is based on the per cent difference between level of energy consumption in the adjustment period in the original baseline and that on the day of DR activation.

Day types and times for which baselines were developed and tested

It is to be expected that DR will seek to bid into the market when prices are relatively high. An analysis was undertaken to determine when high price events tend to occur. The analysis considered the top 100 half-hourly prices that occurred in each NEM region in each year from 2017 through 2019. The analysis revealed that:

- A majority of the high price events (55%) occurred between 3:30 and 8pm, with the highest proportion occurring between 5:30 and 7:00pm[2].
- 50.4% of all price events occur in the first quarter
- 85% of the high price half hours occur on weekdays
- Between 70% and 85% of price events occur in Q1 and Q3; in Queensland the proportion occurring in Q1 is higher than in other states
- In most states approximately 60% to 80% of the high price events occur in two or more contiguous intervals with the balance occurring as single isolated intervals; Tasmania is the exception with most of its high prices occur as single half hours.

Based on these findings the candidate baselines were tested in the following time periods:

Time periods for which the candidate BMs were tested

Time period name	Specific hours
Afternoon/evening	3;30 to 8:00 PM
Short early evening	5:30 to 7:00 PM
Morning	7:00 to 9:00 AM

2 All times used in this report are Australian Eastern Standard Time (AEST).

Customer segments considered

The assessment of the applicability of the candidate BMs was undertaken with respect to three customer segments, defined by annual consumption. The table below shows the three segments and the number of NMIs within each by NEM region.

Customer segment sizes and distribution by NEM region

Segment name (annual consumption)	ACT/NSW	QLD	SA	TAS	VIC	Total NMIs
Medium sized businesses (160 to 750 MWh)	14,780	7,850	3,230	1,340	11,590	38,790
Large commercial & industrial customers (750 MWh - 100 GWh)	5,554	3,829	1,194	501	4,765	15,843
Very large industrial customers (More than 100 GWh)	4	2		2	14	22

The applicability of the candidate BMs was tested for all of the NMIs in Large commercial and industrial and Very large industrial customer segments. Due the large number of NMIs in the Medium-sized business customer segment the analysis was undertaken with a 10% sample on a regional basis.

Metrics used in assessing candidate baselines

The study assessed the outturn accuracy, bias and variability of each of the candidate BMs when applied to each customer segment in each region and in each of the price event windows in each calendar quarter from 2017 through 2019. The metrics used were the same as those used by DNV-KEMA in the 2013 study AEMO commissioned to select a baseline for assessing the performance of DR in the RERT.

Those three metrics were:

- 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 the daily variations.

- Based on the threshold specified in the RERT Panel Agreement and the AEMC Rule change, as discussed in Section 1.2, 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%.

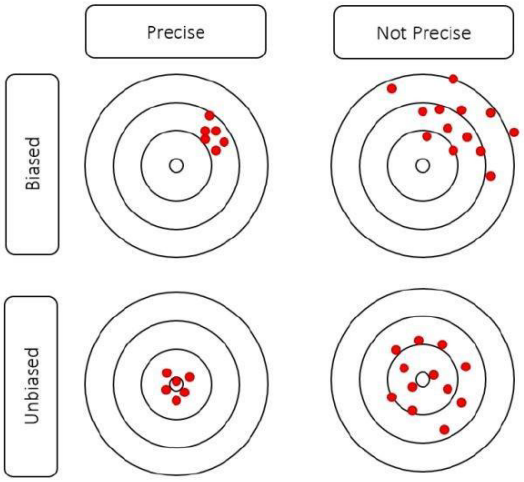
The RERT does not currently use the ARE metric as a specific acceptance criterion, nor is it specifically mentioned in the AEMC Rule change. In this study we have used our interpretation of the results of the study DNV-KEMA undertook for AEMO in 2013, which is that a median ARE in the range of ±0.01 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'[3].

- 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.

  Neither the RERT Panel Agreement nor the AEMC Rule change uses the RER as an acceptance criterion. For this and other reasons[4] the RER was calculated in the analysis undertaken for this study, but it is not used as a criterion in assessing the candidate BMs.

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

The relationship of bias and precision to accuracy



In assessing the results:

- Primary importance was given to the RRMSE score as the measure of the accuracy provided by the BM. The applicability of each candidate baseline was assessed with regard to the median RRMSE it produced in the NMIs within each customer segment in each region and in each of the three event windows in each quarter of 2017, 2018 and 2019.
- Also of importance, but secondarily to accuracy, was the bias associated with each BM. This was assessed with reference to the median ARE score, which was calculated for the NMIs within each customer segment, region and timeframe as the RRMSE.

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

4 See the discussion in Section 3.1.2 of the Functional Specification in Appendix B.

- The RER was also calculated but not explicitly considered as a selection criterion for the candidate BMs[5].

Considerations in selecting the baseline approach to be used

Of primary importance in selecting the BM approach and level of accuracy to be required for participation in the WDRM was the following statement in the AEMC's Final Determination:

> *The Commission considers the metrics produced by AEMO should require baselines to exceed the levels of accuracy considered 'good' in the AEMO-ARENA demand response RERT trials.*[6]

In making this statement, the Commission referred to a document that was produced for ARENA as part of the ARENA-AEMO Demand Response RERT Trial, which stated that "several groups including AEMO, PJM and KEMA have used the 10 per cent and 20 per cent thresholds as the definition of 'good' and 'acceptable' levels of accuracy".[7]

In addition, the Phase 1 analysis sought to strike a balance between the absolute level of accuracy required for participation in the WDRM and the proportion of customers that would be eligible to participate in the WDRM given that accuracy requirement. To do so, an assessment was undertaken of how the proportion of customers eligible to participate in the WDRM would change under each of the candidate BMs using the three different combinations of RRMSE thresholds and the frequency at which a NMI would need to meet that level shown in the table below:

Accuracy and frequency thresholds used to assess potential eligible NMI population size

RRMSE	Frequency of meeting RRMSE threshold[8]
0.1 or less	100%
0.2 or less	100%
0.1 or less	90%

Findings and recommendations

Key findings of the Phase 1 analysis were that the median RRMSEs produced by the candidate BMs:

- Varied materially within each region and to a moderate extent across regions
- Varied to some extent across calendar quarters, with results generally poorest for Q4
- Varied materially across the three event windows that were assessed.

---

5 The RER score generally mirrors the RRMSE and in other studies has not been considered to be of material importance when assessing alternative BMs.

6 AEMC, *Wholesale demand response mechanism, Rule determination,* June 2020, p 188.

7 Oakley Greenwood, *Baselining the ARENA-AEMO Demand Response RERT Trial*, prepared for ARENA, September 2019, p. 7. However, it is important to note that in the case of the WDRM the Rule change requires the accuracy criterion to be applied to each prospective or participating NMI within a DRSP's portfolio. By contrast, in the RERT, where the proponent has multiple sites, the accuracy criterion is applied on the portfolio as a whole (though a single NMI participant in the RERT would be subject to the accuracy requirement). This in and of itself will make any particular accuracy criterion significantly more stringent in the WDRM as compared to the RERT.

8 Baselines for each event window were calculated for every day within each quarter of each year. Meeting the RRMSE threshold required that the NMI had to achieve that score with that frequency for each quarter being considered.

The table below provides information on the 'best' BM in each customer segment based on the results of the Phase 1 analysis, where 'best' is defined as the BM that produces the lowest median RRMSE and a median ARE within the acceptable range[9].

'Best' BM by customer segment

Segment (annual consumption)	Best BM (lowest median RRMSE)	Comments
Medium-sized business customers (160 to 750 MWh)	Multiplicative adjustment with 20% cap	• Produces a median RRMSE in the 'acceptable' range in both the afternoon/evening and short early evening event windows across the 2017-2019 period as a whole • ARE median values are in the 'good' range • A multiplicative adjustment with a 40% adjustment cap produces marginally better median RRMSEs in Q1 in all years • The additive adjustment with a 20% cap produces the lowest RRMSEs in the morning event window
Large commercial and industrial customers (750 MWh to 100 GWh)	Multiplicative adjustment with 40% cap	• The multiplicative adjustment with a 20% cap provides 'good' median RRMSE in both the afternoon/evening and short early evening event windows across the 2017-2019 period as a whole • ARE median values are in the 'good' range • However, the multiplicative adjustment with a 40% adjustment cap produces marginally better median RRMSEs in Q1 in all years • The additive adjustment with a 40% cap produces the lowest RRMSEs in the morning event window
Very large industrial customers (Over 100 GWh)	None – recommendation is to consider allowing site-specific BMs	• Analysis was undertaken at the NMI level rather than across NMIs due to the small number of customers in this segment • Five of the 21 NMIs in this segment showed nett export during some, and in some cases, all quarters, which eliminated them from further consideration • In each of the other 16 NMIs, at least one of the BMs provided median RRMSE scores in the 'good' range in each of the event windows (with he exception of one NMI in the afternoon/evening event window) • But none of the BMs proved to be 'best' across a significant number of the 16 sites • Several sites were able to provide only 'acceptable' median ARE scores in each of the event windows

As noted in the table above, the best' BM differs by customer segment and time of day:

- For both Medium-Sized Business Customers and Large Commercial and Industrial Customers, a multiplicative adjustment approach was shown to be better in the afternoon/evening and short early evening event windows, while an additive approach proved better in the morning event window.
- But a different adjustment cap was seen to perform better in those two customer segments. A 20% cap proved better for Medium-Sized Business Customers while the 40% cap was better for Large Commercial and Industrial Customers.

9 The test for the 'best' BM was undertaken on a quarter by quarter and an annual basis and across the full 2017-2019 timeframe for each customer segment in each NEM region.

- No single BM was shown to consistently provide the best results in among Very Large Industrial Customers.

However, from an administrative perspective:

- Ideally, the threshold level of accuracy and frequency to be applied should not vary by region, segment or season. Rather the selection of these thresholds should reflect a balance between what is deemed to be an acceptable level of error (bounded by the accuracy requirements laid out in the Rule change) and a reasonable level of customer eligibility.
- In particular, the use of the same BM in all regions and for all customer segments is likely to be preferable for the perceived equity of the WDRM as well as for administrative simplicity. However, where different BMs can be expected to be able to produce a higher level of eligibility at the selected level of accuracy when applied to different regions or customer segments, a case could be made for the applicable BM to vary. From an administrative perspective such a variation would probably be easier to implement regionally than by customer segment. Variation by season or event window would likely be very difficult operationally and is not recommended.

Based on these considerations, if AEMO would prefer to use same BM in all regions and for all customer segments, our recommendation would be to:

- Use the multiplicative adjustment BM with a 40% cap on the adjustment for both the Medium-Sized Business Customer and the Large Commercial and Industrial Customer segments in all time periods in all seasons and NEM regions. Our reasons are as follow:
  - The multiplicative approach is preferred because the afternoon and evening event windows include more high price events than any other parts of the day, including the morning event window
  - The 40% adjustment cap proved better for the Large Commercial and Industrial Customer segment which can be expected to have greater DR potential than the Medium-Sized Business Customer segment.
- Consider allowing Very Large Industrial Customers that want to participate in the WDRM to use one of the other pre-period only adjustment BMs where that BM provides a lower median RRMSE, particularly where that alternative will provide a median RRMSE that qualifies the site for participation when the multiplicative BM with 40% adjustment cap would not do so. We note that this would:
  - Increase the amount of DR that could be offered into the WDRM
  - Provide the first step in considering an expansion of the BMs used in the WDRM.

This approach, we believe, would provide a BM for the WDRM that:

- Is simple for customers and DRSPs to understand
- Reduces administrative burden for customers, DRSPs and AEMO
- Maintains a proper focus on accuracy in line with the AEMC's Rule change, but balances that with both simplicity and the desire to provide a reasonable level of eligibility for customers
- Provides the opportunity for further evolution and sophistication in the BMs used in the WDRM.

1. Introduction

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[10]. DRSPs will be able to engage directly with customers and 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 mechanism will provide greater opportunities 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 price.

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 registration and compliance testing of candidate WDRUs and the settlement of WDRUs and DRSP portfolios in DR events.

1.2. Purpose

This objective of this project is to:

- Test the applicability of a specified set of BMs to the loads of the large business customers that were defined as being eligible to seek to register as WDRUs. The BMs that have been specified are variations of the CAISO 10 of 10 BM that were analysed for AEMO by DNV-KEMA when the original Demand Response Mechanism rule change proposal was being considered.[11]
- Provide recommendations regarding:
  - The most appropriate threshold values for the accuracy/bias metrics to be used with the BM determined to be best suited to the customer base (or possibly a particular defined segment within the customer base), and

10 In doing so, the DR that is bid will be treated essentially as a scheduled load.

11 DNV-KEMA, *Development of Demand Response Mechanism Baseline Consumption Methodology - Phase1 and 2*. The final report and results were published in October 2013,

- The details of how the BM is to be implemented by AEMO with regard to specific WDRUs, including the data required, the statistical tests to be undertaken, any WDRU-specific considerations that will need to be able to be addressed in the implementation process, and the complexity and cost the procedure can be expected to impose on both AEMO and the candidate WDRU.

1.3. Overview of scope and approach

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 was 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.

  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,[12] 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.

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

Appendix A: contains a copy of AEMO's Brief for the project. It should be noted that the Brief asked for the Phase 1 analysis to be completed in two weeks and that the work for both Phases be undertaken on a fixed fee basis.

1.4. Caveats and limitations

It should be noted that:

- The candidate BMs analysed in this study was limited to a set of variants of the CAISO 10 of 10 methodology
- Discussions of the proportion of customers eligible to participate in the WDRM are maximums – they are the greatest number of NMIs within a particular segment region and timeframe who could be expected, based on historical consumption data, to meet a specific accuracy threshold with a given frequency. Actual participation rates will vary because (a) not all customers who could be eligible will choose to participate, and (b) a customers' future consumption (i.e., when seeking to register for the WDRM) may differ from that of the past.

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

1.5. Organisation of this report

The remainder of this report is organised as follows:

- Section 2 provides a detailed description of the approach taken in the Phase 1 analysis
- Section 3 presents the results of the analysis undertaken in Phase 1 of this study, and
- Section 4 discusses the implications of the results of the analysis and the recommendations that flow from them.

1.6. Next steps

Phase 2 of the study will provide information on how the BM selected based on the results of the Phase 1 analysis can be adapted for use in the implementation of the WDRM. Specifically, Phase 2 will provide recommendations regarding:

- The most appropriate threshold values for the accuracy and bias metrics to be used in administering the WRDM, and
- A methodology for calculating the key accuracy/bias metrics that can be used in the assessment of the eligibility of individual NMIs for participation in the WDRM (i.e., a methodology for determining their ability to meet the accuracy and bias thresholds required for participation)

2. Approach

The approach that we proposed in response to AEMO's Brief was further refined in consultation with AEMO in response to issues and needs that arose once the project began. Appendix B: provides a copy of the Functional Specification that we developed in consultation with AEMO in response to those evolving issues.

2.1. Objective of the Phase 1 analysis

As stated in the project Brief, the objective of the Phase 1 analysis was to "test the efficacy of a variety of 'RERT like' BMs with commonly accepted adjustment approaches. The analysis is to use multiple metrics and recent NEM data for a range of potential WDRM participants". More specifically, the objective of the Phase 1 analysis was to identify which of the candidate BMs could be expected to provide the most accurate assessment of the DR delivered by WDRUs in the settlement of actual DR events.

The remainder of this section describes how we addressed this objective, but it should be noted that the Brief did not specify a threshold test for 'efficacy'. In its *Wholesale demand response mechanism, Rule determination* published in June 2020, the Australian Energy Market Commission (AEMC) stated:

> *The Commission considers the metrics produced by AEMO should require baselines to exceed the levels of accuracy considered 'good' in the AEMO-ARENA demand response RERT trials.*[13]

In making this statement, the Commission referred to a document that was produced for ARENA as part of the ARENA-AEMO Demand Response RERT Trial, which stated that "several groups including AEMO, PJM and KEMA have used the 10 per cent and 20 per cent thresholds as the definition of 'good' and 'acceptable' levels of accuracy".[14]

In particular, in its study for AEMO, KEMA stated that a Relative Root Mean Square Error (RRMSE) less than 10% should be seen as the appropriate acceptance criteria for a successful BM, and that was the rationale for selecting the CAISO 10 of 10 BM with an additive adjustment as the methodology for the RERT.

However, it should also be noted that the RERT Panel Agreement itself makes reference to an accuracy threshold of 20 per cent, as measured by the Relative Root Mean Square Error (RRMSE) statistic.[15]

---

13 AEMC, *Wholesale demand response mechanism, Rule determination,* June 2020, p 188.

14 Oakley Greenwood, *Baselining the ARENA-AEMO Demand Response RERT Trial*, prepared for ARENA, September 2019, p. 7. However, it is important to note that in the case of the WDRM the Rule change requires the accuracy criterion to be applied to each prospective or participating NMI within a DRSP's portfolio. By contrast, in the RERT, where the proponent has multiple sites, the accuracy criterion is applied on the portfolio as a whole (though a single NMI participant in the RERT would be subject to the accuracy requirement). This in and of itself will make any particular accuracy criterion significantly more stringent in the WDRM as compared to the RERT.

15 The RERT Panel Agreement for 2019-20 (which is the version that was operant at the time the AEMC issued its Rule Determination, states that "AEMO may measure the accuracy of the unadjusted baseline by determining the unadjusted baseline's relative root mean squared error (RRMSE) by comparing the Reserve Provider's unadjusted baseline against the 60 days not being Activated Days immediately preceding the weekday on which reserve was activated and for which the calculation is being made and if they vary from each other by a value greater than or equal to 20%, AEMO may adjust the variables which are used to determine the unadjusted baseline to ones which AEMO determines, acting reasonably, more accurately reflects the Reserve Provider's typical demand" (page 48). That same clause appeared in the prior and subsequent years' Panel Agreements. Where an adjustment was undertaken under that provision the adjustment itself was to be a symmetrical, additive adjustment limited to 20% of the reserve capacity in the event that the adjustment was determined to be a positive factor.

Phase 1 also included consideration of the impact that the threshold set for the BM would have on the potential proportion of customers that would likely be eligible to participate in the WDRM. While the level of potential eligibility was not deemed to be as or more important that the accuracy with which the quantum of demand response would be able to be measured, it was recognised that this proportion would be of interest to various stakeholder groups, and in the event that the number of potential providers of demand response that can be measured with a suitable level of accuracy was determined to be exceedingly small, the administrative cost of providing the WDRM might exceed the benefits to be derived from it.

Reflecting these concerns, at a practical level, the objective of Phase 1 as undertaken was to:

- Assess the accuracy of each of the candidate BMs as defined by the median, 10th and 90th percentile RRMSEs
- Identify the trade-offs between the candidate BMs in terms of their potential s to provide a reasonable level of eligibility among customers at an agreed accuracy threshold while also minimising the potential for gaming
- Identify procedures that can be used to limit the potential for inaccurate baselines at the NMI level.

2.2. BM variants assessed

2.2.1. Description of the BMs

AEMO's Brief specified testing of the following BMs:

- CAISO 10 of 10 - no adjustment
- CAISO 10 of 10 - pre-period additive adjustment with no cap
- CAISO 10 of 10 - pre-period additive adjustment with 20% cap
- CAISO 10 of 10 - pre-period additive adjustment with 40% cap
- CAISO 10 of 10 - pre-period multiplicative adjustment with no cap
- CAISO 10 of 10 - pre-period multiplicative adjustment with 20% cap
- CAISO 10 of 10 - pre-period multiplicative adjustment with 40% cap

In discussion with AEMO, the 'no adjustment' and 'no cap' options were not pursued in Phase 1 as the value of these approaches was considered to be low.

The use of a pre-period adjustment refers to a methodology for adjusting the consumption level of the baseline to reflect the consumption level of the NMI in the time period preceding DR activation on the price event day. The methodology increases (or decreases) the consumption level of the NMI's baseline based on the difference between the average consumption in the six trading intervals that end one hour prior to commencement of the DR activation on the day for which the BM is being constructed and the average consumption in those trading intervals on the days included in the BM.

In an additive adjustment approach the difference is applied in absolute terms (i.e., the number of kWh added or subtracted from the baseline) up to a cap (i.e., the adjustment is not allowed to exceed some specified percentage of the difference between the adjustment window consumption in the baseline (before adjustment) and on the price event day.

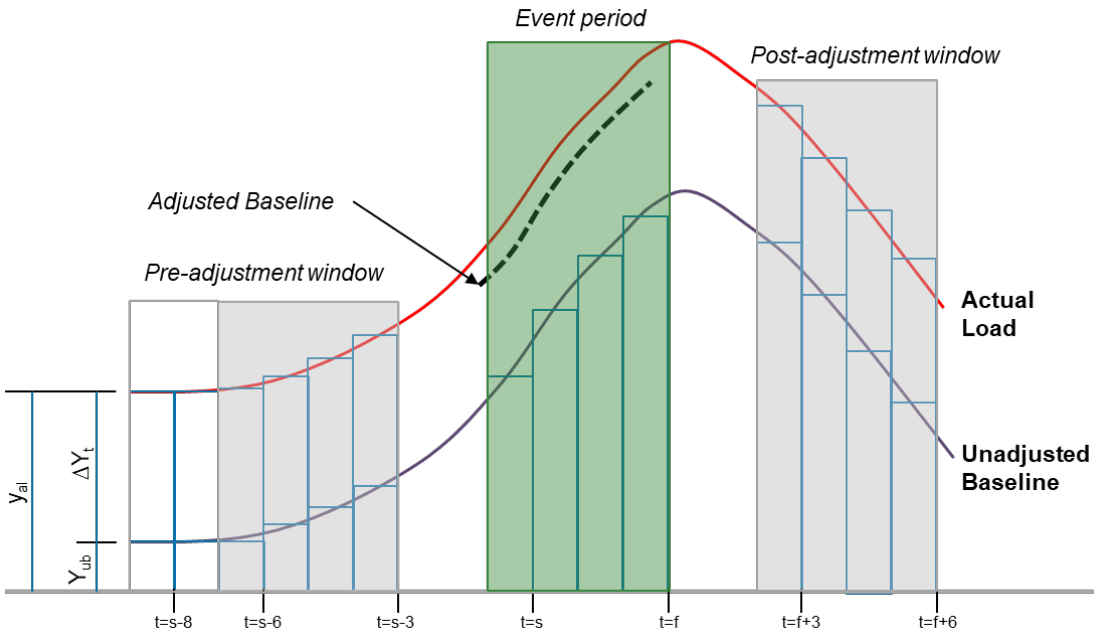
In a multiplicative adjustment approach the difference between the adjustment window consumption in the (unadjusted) baseline and on the price event day is expressed and applied as a percentage of the baseline (either upward or downward) but is still subject to the specified cap on the adjustment. Based on our experience in testing the accuracy, bias and variability of a number of other BMs OGW suggested that the testing include BMs that employ a post-event adjustment period in addition to the pre-event adjustment. AEMO accepted this suggestion and the following BMs were added to those above for testing in the project:

- CAISO 10 of 10 - pre/post period additive adjustment with 20% cap
- CAISO 10 of 10 - pre/post period additive adjustment with 40% cap
- CAISO 10 of 10 - pre/post period multiplicative adjustment with 20% cap
- CAISO 10 of 10 - pre/post period multiplicative adjustment with 40% cap

When pre- and post-period adjustments are both used, an adjustment window after the end of the DR activation is also applied for amending the baseline consumption level at the end of the activation period based on the observed difference between the NMI's consumption on the baseline days as compared to the activation day.

Figure 1 shows the baseline prior to adjustment and how the baseline is adjusted in light of the metered load during the adjustment window on a day on which DR is activated.

Figure 1: Pre- and post-period adjustments to the baseline



2.3. Day types and trading intervals analysed

2.3.1. Eligible days

The Brief requested a recommendation on the number of eligible days to be used in the testing of the candidate BMs. Our interpretation was that 'eligible days' refer to what is called the 'baseline window' in the RERT Panel Agreement.

In accordance with the RERT Panel Agreement and given the fact that there are unlikely to be a significant number of days other than weekends and public holidays in the consumption data being used in the analysis of the candidate BMs that are not qualifying days[16], it was decided to keep the number and description of eligible days (baseline window) the same as in the RERT Panel Agreement: 45 calendar days.

2.3.2. Day types

The Brief requested that recommendations be made on the number and types of days to be used in the BMs to be tested. The approach taken in case is detailed below.

- The number of candidate days to be used – In accordance with the nature of the BMs specified for testing it was decided that the number of candidate days to be used in the Phase 1 analysis would be the same as specified in the CAISO 10 of 10 methodology as used in the RERT. That is, each baseline was developed from the consumption data for that NMI in the prior 10 weekdays, in the case where the baseline was being developed for a weekday.

  It was noted that the RERT Panel Agreement does not specify how baselines are to be formulated for weekend days. We recommended and AEMO accepted adopting the CAISO 10 of 10 approach for developing baselines for weekend days. That approach develops the baseline from the average consumption on the four most recent weekend days out of the four most recent weekend.

- Candidate day conditions to be considered – The Brief sought recommendations on how public holidays, zero consumption days and negative consumption periods within days should be handled in the analysis of the candidate BMs. The decisions made in each case were as follow:

  - Public holidays – The public holidays relevant to each jurisdiction were excluded from the analysis because it could reasonably be assumed that their loads will be characterised by different consumption patterns from the norm for that site.

  - Zero consumption days – In accordance with the fact that there are no provisions in the RERT Panel Agreement to exclude days on the basis of load type except for weekends, public holidays or activation days from the calculation of a baseline, zero consumption days were included in the analysis of the candidate BMs.

    Flags were used to identify days for which the baseline included zero consumption days (or zero-consumption event periods). These flags allowed determination of the frequency at which these types of days/periods occur and the degree to which they affect the predictive power of the baseline

  - Negative consumption periods within days – Negative consumption intervals were assumed to represent net exported electricity. Negative consumption intervals were flagged but included in the baseline calculations and analysis. As above, these flags can be used to determine the frequency at which these types of periods occur and the degree to which they affect the predictive power of the baseline.

16 That is, there were no activation days in the consumption dataset used in the analysis.

2.3.3. Trading intervals used in the BM tests

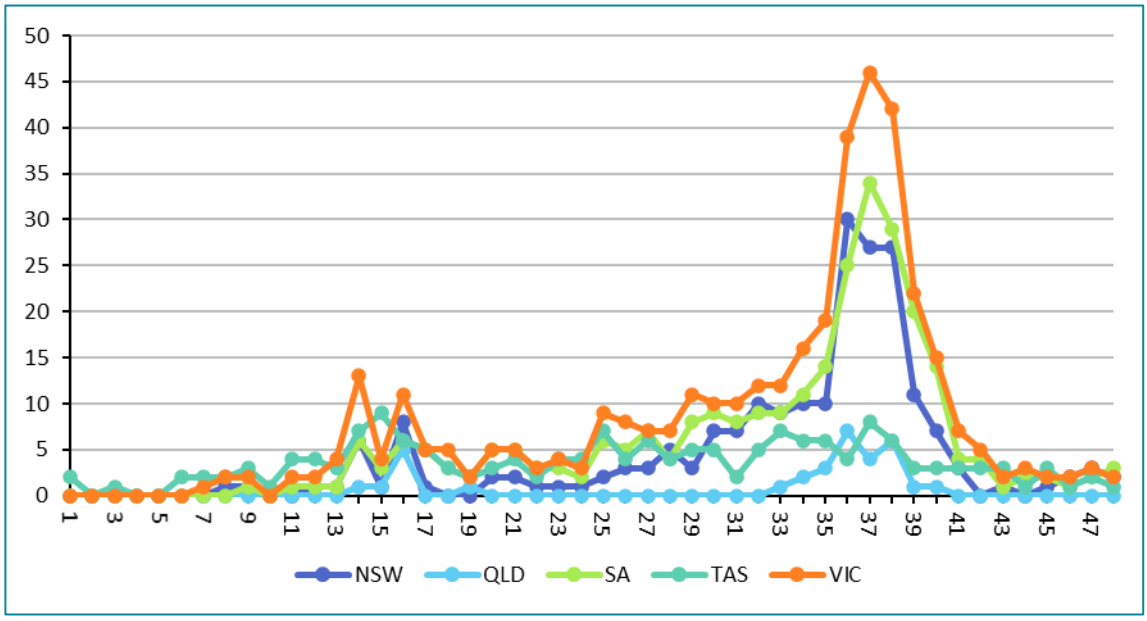
The work conducted for AEMO by DNV-KEMA in 2013 for testing different baseline methodologies for use in the RERT identified that the top 100 half-hour demands for all regions occurred between 1pm and 5pm and between 6pm and 7pm in the months of October to March. Based on that, DNV-KEMA used 1pm to 7pm as the event period for the calculation of their baseline test statistics.

Activation for the purposes of the WDRM will be based on wholesale price levels rather than demand levels (though the two are often closely related). As a result, we undertook an analysis of the time, day of the week and seasonal distribution of high price events. The analysis assessed approximately the top 100 price intervals in each of the NEM regions in each of the years 2017, 2018 and 2019.

Interestingly, the top (approximately) 100 price events in each jurisdiction exhibit prices of approximately $300/MWh and above. It is worth noting that this corresponds to the cost at which caps can generally be purchased in the market, which constitutes the natural competition for DR. It is also likely to be the threshold at which demand response might begin to be activated under the WDRM.

To provide further insight into the intervals and time of week and year where the probability of a high price event might occur, a fixed threshold of $270/MWh was set across all jurisdictions. A majority of these price events (55%) occurred during the intervals 32 to 40 (i.e., the half hours ending from 4pm to 8pm). It was also found that 50.4% of all price events fall in the first quarter. Figure 1 shows the weighting towards the afternoon high price events with a large portion in the trading intervals 36 to 38 (i.e., those occurring in the three half hours ending between 6pm to 7pm).

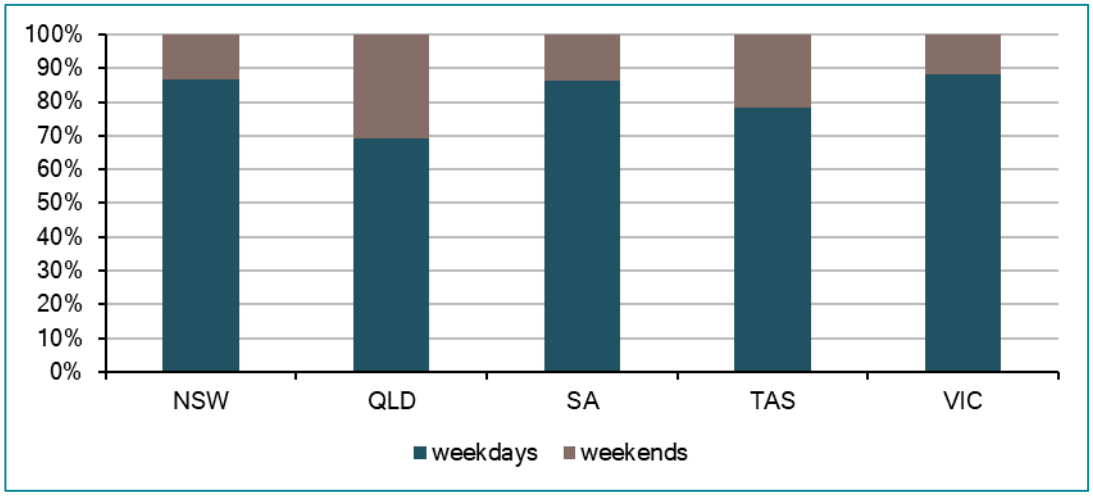
Figure 2: Annual accumulative half hour high price intervals all days for 2019.



OGW Analysis

Figure 1 shows the split between weekdays and weekends that high price events occur across the years 2017 to 2019 for all high price intervals. Most states show a higher weighting to weekday price intervals with only approximately 15% occurring during the weekends. Queensland shows a higher proportion to weekend price events with 30% occurring on weekends; however, on closer examination most of these weekend events were isolated to Jan and Feb in 2017. For the years 2018 and 2019 the distribution was like the other states.

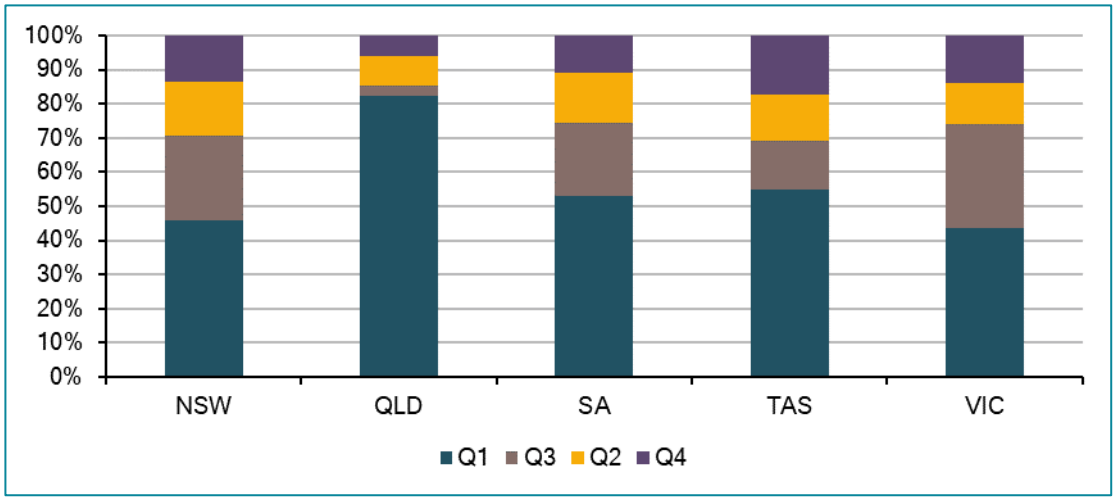
Figure 3: Breakdown of eligible price intervals occurring on weekdays vs weekends



OGW analysis

Figure 2 shows the applicable high price intervals and how they occur across yearly quarters for all price intervals during the tears 2017 to 2019. For all states, the 70%-85% of price events occur in Q1 and Q3. Queensland is slightly different where a large majority occur in Q1 compared to the other states.

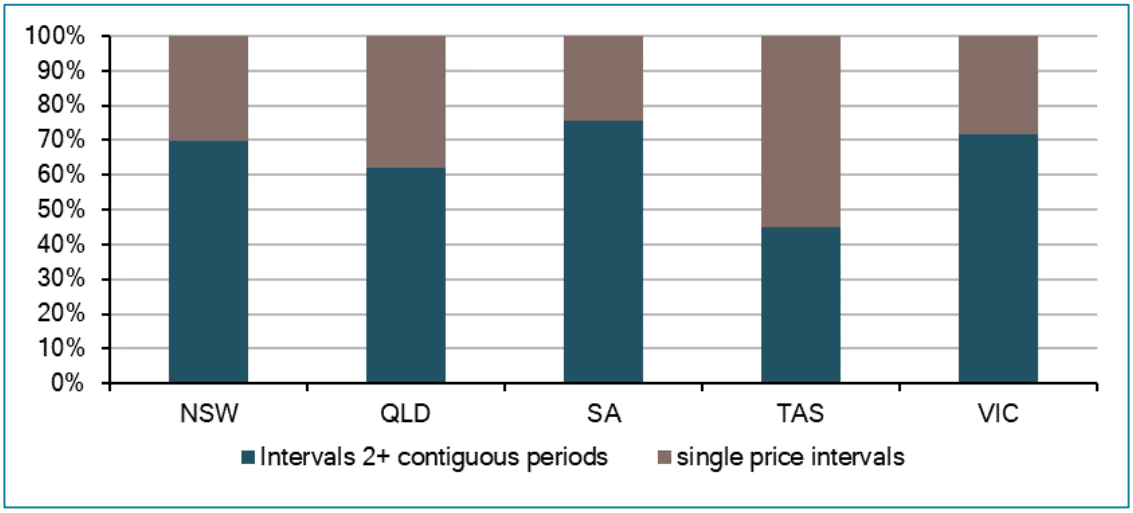
Figure 4: Breakdown of applicble high price intervals occurring during annual quarters



OGW analysis

Figure 3 provides a summary of the types of price intervals that occur. For all states except Tasmania the high price intervals occur approximately 60%-80% of the time in two or more contiguous intervals with the balance occurring as single isolated intervals. Tasmania is the opposite where most of the price intervals are single events.

Figure 5: Breakdown of eligible price intervals that occur during contiguous price intervals compared to singular price intervals



OGW analysis

The price event analysis revealed that:

- A greater weighting to weekdays compared to weekends.
- Most price events occur in the q! and Q3 yearly quarters with a large number occurring in January.
- Events tend to occur during consecutive half hours over a couple of days during trading intervals 30 to 40 (i.e., the half hours from 2:30 to 8:00pm) with a high tendency to occur in intervals 36 through 38 (i.e., the half hours from 5:30 to 7:00pm)
- There are some single half-hour event periods scattered much more broadly throughout the year, but there is some question about the ability of WDRM aggregators to forecast their occurrence and whether WDRM participants would consider a half-hour worth the effort of a response.

Based on the price event analysis the trading intervals shown in Table 1 below were agreed to be used in the BM testing.

Table 1: Trading intervals used in the BM testing

Test no	Trading intervals included	Half hours included
Test 01 - morning peak	Trading intervals 15 through 18	Half hours ending at 07:30 through- 09:00
Test 02 - evening peak	Trading intervals 36 through 38	Half hours ending at 18:00 through 19:00
Test 04 - afternoon high price	Trading intervals 32 through 40	Half hours ending at 16:00 through 20:00

Appendix C: contains further details from the price event analysis.

2.4. Segmentation and sampling approach

2.4.1. Description of segments

In consultation with AEMO it was decided to assess the applicability of the candidate BMs on a segmented basis. The segmentation was based on annual consumption at the NMI level using the following thresholds:

- More than 100 GWh – these are very large industrial customers
- More than 750 MWh/pa to 100 GWh – these are large commercial and industrial customers
- More than 160 to 750 MWh[17] -- these are medium-sized business customers many of which are likely to be commercial rather than industrial facilities.

AEMO provided consumption records for all NMIs within these consumption bands for the period 1 March 2016 through 31 August 2020. Table 2 shows the number of NMI consumption records[18] received from AEMO.

Table 2: Number of NMI consumption records received within each segment, by jurisdiction

Annual consumption range	ACT/NSW	QLD	SA	TAS	VIC	Total NMI counts
G: 100+ GWh	4	2		2	14	22
F: 0.75+ GWh to 100 GWh	5554	3829	1194	501	4765	15843
E: 160+ MWh to 750 MWh	14780	7850	3230	1340	11590	38790

2.4.2. Overview of the sample analysed

In undertaking the analysis:

- The analysis was undertaken on a NEM region basis, which meant that the ACT NMIs were combined with those of NSW
- A 10% random sample of the smallest consumption segment was used in the analysis due to the large number of NMIs in that segment.

2.4.3. Independent variables considered

The AEMO Brief asked that consideration be given to the use of relevant independent variables as part of the BM testing and listed industry type, site characteristics, weather as candidates.

In discussion it was noted that AEMO did not have any information on either the industry type or site characteristics (other than location) of the NMIs in its database, and acquiring such information far exceeded the time and budget available for the study. In addition, these factors do not figure in any of the BMs to be tested.

17 Customers with annual consumption below 160 MWhpa were not included in the analysis as (a) it is not clear that they are eligible to participate in the first year of the WDRM, and (b) the large number of customers within this segment posed complexities that exceeded the timeframe and resources available for this study.

18 To comply with privacy requirements, OGW generated an alias for each of the NMIs and a file that could be used to map an alias back to its associated NMI. The mapping file and the original datasets were then given back to AEMO; OGW did not retain a copy of either the original data or the mapping file. AEMO can retain the map and the original data set for its own purposes and for the duration it deems necessary.

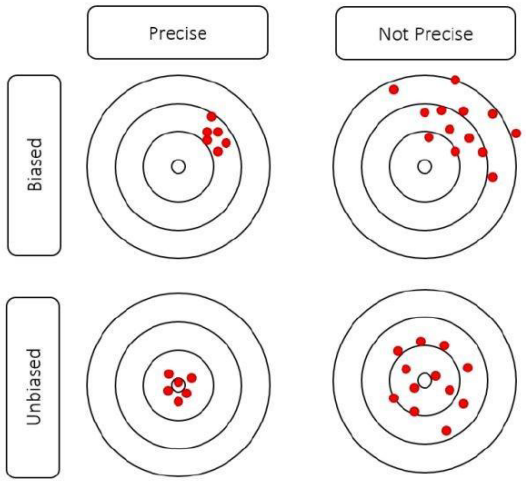
In previous work on BMs for ARENA, OGW has identified that the use of weather variables can improve the accuracy of BMs for weather-sensitive loads. However, weather variables are not included in any of the BMs that were to be tested under the scope of work of this project. In discussions, AEMO noted that while there is evidence that the inclusion of weather variables can increase the pool of end-use facilities whose demand response can be assessed with a suitable level of accuracy, it entails a level of additional complexity which could not be accommodated within the timeframe available for readying the WDRM for implementation[19]. AEMO expects the BMs available for use in the WDRM to evolve over time and the issue of weather-sensitivity is an area of obvious interest.

2.5. Metrics applied

2.5.1. Overview of the metrics used

The metrics used in the study assess the candidate BMs in terms of their accuracy, bias and precision. Figure 4 below provides a graphic representation of the differences between these three attributes.

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



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 the daily variations.

  Based on the threshold specified in the RERT Panel Agreement and the AEMC Rule change, as discussed in Section 1.2, 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.

19 Two examples of the complexities introduced by the inclusion of weather variables are (a) the potential importance of the proximity of the NMI to the weather station whose data is to be used, and the corresponding need to geo-locate each NMI.

- 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, nor is it specifically mentioned in the AEMC Rule change. In this study we have used our interpretation of the results of the study DNV-KEMA undertook for AEMO in 2013, which is 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].

- 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.

  Neither the RERT Panel Agreement nor the AEMC Rule change uses the RER as an acceptance criterion. For this and other reasons[21] the RER was calculated in the analysis undertaken for this study, but it is not used as a criterion in assessing the candidate BMs.

Further information on the mathematical formulations and the use of these three metrics is provided in the Functional Specification presented in Appendix B.

2.5.2. Use of the metrics in decision-making

Because the mathematical formulation of the RRMSE combines the systematic errors measured by the ARE (bias) and the variability of errors captured by the RER (variability), it is given primary emphasis in assessing BMs in this and most other studies.

Because our analysis was undertaken at the calendar quarter level, we calculated the RRMSE for each NMI for every weekday in each quarter. This allows us to calculate median RRMSE for all NMIs in each segment within each region for each BM by calendar quarter and the range of those medians across the various BMs. It also allows us to assess the average RRMSE of any particular NMI in any quarter as well as the frequency with which any particular NMI achieves or exceeds any specific threshold RRMSE level. The ARE is next in importance as it provides a measure of the direction of the net error in the baseline.

The RER score generally mirrors the RRMSE and is not considered to be of material importance when assessing alternative BMs. This has been borne out in the analysis.

2.6. Description of analytic procedures and decision thresholds employed

2.6.1. Analytic procedures and decision thresholds employed

Two sets of analyses were undertaken:

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

21 See the discussion in Section 3.1.2 of the Functional Specification in Appendix B.

- An assessment of the accuracy of the baselines produced through the use of each of the candidate BMs. This assessment examined the medians of the three metrics described above across the candidate BM approaches, noting the best performer against the RRMSE metric, and
- An assessment of how the proportion of customers eligible to participate in the WDRM changes under each of the candidate BMs using three different combinations of RRMSE thresholds and the frequency at which a NMI needs to meet that level. Those combinations are shown in Table 3 below.

Table 3: Accuracy and frequency thresholds used to assess potential eligible NMI population size

RRMSE	Frequency of meeting RRMSE threshold[22]
0.1 or less	100%
0.2 or less	100%
0.1 or less	90%

- The 0.1 RRMSE with 100% frequency was selected to reflect a level of 'good' accuracy[23] (±10%) being achieved in all of the site's baselines considered over the timeframe in question. Such a threshold would provide a very high level of confidence to AEMO about the accuracy of the chosen BM and the level of error likely to be associated with its use.
- The 0.2 RRMSE with 100% frequency was used to assess how the proportion of eligible NMIs could be increased if AEMO were to use a BM that provides an 'acceptable' level of accuracy[24]. The 100% frequency threshold would still provide a high level of confidence to AEMO about the level of error likely to be associated with the use of the chosen BM.
- The 0.1 RRMSE with 90% frequency was used to assess the impact of the use of a 'good' accuracy level but not requiring it to be met in all of the site's baselines over the timeframe in question. The use of this combination of RRMSE and frequency provided a means for assessing the potential trade-off between the level of accuracy provided by the chosen BM and the level of confidence provided about the level of error likely to be associated with its use. It is important to note that the level of error that would occur in the remaining 10% of the cases would not be known and as a consequence the total quantum of error possible under this arrangement was not able to be assessed at the level of analysis undertaken in Phase 1.[25]

The analysis was capable of significant granularity. It provided over 180 cells (5 regions, 3 segments and more than12 calendar quarters) in which the accuracy, bias, variability and eligibility potential of each of the 8 BMs in each of the 3 event windows could be assessed.

22 Baselines for each event window were calculated for every day within each quarter of each year. Meeting the RRMSE threshold required that the NMI had to achieve that score with that frequency for each quarter being considered.

23 As defined in a number of studies concerning baselines including work undertaken by KEMA for AEMO in assessing BLMs for the RERT and by OGW in assessing BLMs for ARENA in its SN DR RERT Trial program.

24 This level of accuracy was deemed 'acceptable' in the ARENA-AEMO SN RERT DR Trial program.

25 It should be noted that the level of error in the remaining 10% of the baselines would not necessarily be known in this case. Ways in which this error could be bounded will be considered in Phase 2 of this study.

It must also be recognised, however, that the WDRM is to be implemented across the NEM and that the implementation of any policy requires a balance between accuracy, complexity and administrative cost. Therefore, the analysis sought to determine:

- Which BM could be considered as producing the 'best' level of accuracy when applied in each segment within each NEM region in each quarter for each of the event windows, and whether the selection of a single BM entailed a significant level of compromise in specific regions, segments seasons or event windows.
- The impact that the use of a selected set of RRMSE accuracy and frequency thresholds would have on the proportion of NMIs within each segment that would be eligible to participate in the WDRM in each region, season and event window.

Both of these assessments are described in further detail below.

2.6.2. For the assessment of candidate BMs

The assessment of the accuracy produced by the different candidate BMs was undertaken in the first instance at the segment level. For each of the three segments, the median, 10th and 90th percentile RRMSE, ARE and RER results were calculated for each BM. This calculation was undertaken in the following steps:

- Across all NMIs in each of the five NEM regions,
- For each of the three price event windows of interest, and
- In each quarter of each of the three years 2017, 2018 and 2019.

In each case, the median and range of the RRMSE scores of each of the candidate BMs was reviewed. The BM with the lowest median RRMSE score was preferred as long as its range of RRMSE scores was also acceptable.

The ARE of the preferred BM was also reviewed to ensure it was within the 'acceptable' range[26].

2.6.3. For the assessment of the implications for eligibility

The assessment of the implications of the choice of different combinations of BMs and RRMSE scores was assessed by calculating the proportion of NMIs whose baselines over a specified period of time would meet a specified RRMSE accuracy threshold with a specified frequency for each of the candidate BMs. As a specific example, what proportion of the NMIs within a given customer segment could be expected to have their baselines achieve an RRMSE of 0.1 or less 90% of the time in the first quarter of the year?

As in the case of the assessment of the BMs this analysis was undertaken:

- For the NMIs within each of the three segments
- In each NEM of the five regions
- For each of the three price event windows of interest,
- In each quarter of each of the three years 2017, 2018 and 2019, and
- For each of the candidate BMs.

26 As noted in Section 2.5.1 above, a median ARE in the range of ±0.01 or 1% is considered 'acceptable', and a median ARE in the range of ±0.05 or 0.5% is considered 'good'.

2.6.4. How the results of the analyses are presented

The results of the analyses described above are presented in the following sections of the report. The analyses were undertaken separately for each of the three customer segments described in section 2.4.1 above. This segment-based approach was chosen because there was the possibility that different BMs might be more or less suitable for different customer segments. Within each segment, results are discussed by event window, region and season.

Overall, it was assumed that, from an administrative perspective:

- Ideally, the threshold level of accuracy and frequency to be applied should not vary by region, segment or season. Rather the selection of these thresholds should reflect a balance between what is deemed to be an acceptable level of error (bounded by the accuracy requirements laid out in the Rule change) and a reasonable level of customer eligibility.
- In particular, the use of the same BM in all regions and for all customer segments is likely to be preferable for the perceived equity of the WDRM as well as for administrative simplicity. However, where different BMs can be expected to be able to produce a higher level of eligibility at the selected level of accuracy when applied to different regions or customer segments, a case could be made for the applicable BM to vary. From an administrative perspective such a variation would be easier to implement regionally than by customer segment. Variation by season or event window could be very difficult operationally and is not recommended.

As noted above, the analysis was undertaken on two levels. The first sought to identify the preferred BM, where 'preferred' was defined as the BM that yielded the lowest median RRMSE and an acceptable range of RRMSEs and AREs. The second assessed the trade-off between accuracy, eligibility and the ability to bound likely total error in the choice of BM and how it is applied to individual NMIs.

A specific set of tables was used to present the results of each of these two levels of analysis. The tables used in each level of the analysis are described in the boxes below.

Identifying the preferred BM

These tables assess the accuracy, bias and variability of each of the candidate BMs and were produced by customer segment for each of the three event windows. Each of those tables provides the following information:

- The first three columns state the regions, years and quarters for which results are described in the corresponding row
- The Outcomes column provides summary information regarding the accuracy, bias and variability results across all of the 8 BMs, with the corresponding information for the BM specified in the RERT provided for comparison
- The Comments column identifies the BM that provides the lowest median RRMSE for the regions and timeframes considered and other aspects about that BM in comparison to the other BMs
- The final two columns provide information on the median RRMSE and the type and level of adjustment factor found to be best among the pre-only and pre/post adjustment BMs. These columns also show the median ARE value for those BMs.

Assessing the trade-off between accuracy, eligibility and the ability to bound the expected level of error

These tables show the how the proportion of eligible NMIs within a given customer segment is likely to change as the threshold median RRMSE and the frequency at which a NMI would be expected to meet that level changes for each of the candidate BMs.

Results are presented by customer segment for each of the three combinations of median RRMSE and frequency in each of the three event windows for each NEM region.

3. Results for Medium-Sized Business Customers

This section of the report provides the results for Medium-Sized Business Customers – those customers with annual electricity consumption ranging from 160 MWh to 750 MWh[27]. The majority of these customers will be commercial enterprises rather industrial facilities.

The results of the two levels of analysis are discussed for each of the event windows identified as containing high prices. The extended afternoon/evening event window (which runs from 3:30pm to 8:00pm) is discussed first as it contains the largest proportion of high-price trading intervals annually. Subsequent sections provide parallel discussions of the short early evening event window (which runs from 5:30pm to 7:00pm) and the morning event window (which runs from 7:00am to 9:00am).

Within each event window the results of the analysis are provided as follows:

- First the results of the assessment of the accuracy, bias and variability of each of the various BMs when applied to this customer segment are discussed
- Then the results are discussed of the analysis of the percentage of the NMIs within this segment that would be expected to be eligible to participate in the WDRM under the three different combinations of threshold RRMSE and the frequency with which the NMI met that threshold in the preceding calendar quarter. It should be noted that the analysis of eligibility is discussed only for the afternoon/evening price event window in order to reduce the length of this section of the report. Appendix D provides links to spreadsheets that provide the eligihle percentage of each customer segments in each NEM region for each of the three eligibility thresholds analysed.

Details of the contents of the tables used in the two analyses are provided in section 2.6.4 above.

3.1. Afternoon/evening event window (3:30-8:00pm)

3.1.1. Assessment of accuracy, bias and variability of BMs for Medium-Sized Business Customers

Table 4 below provides the results of this analysis for all 8 candidate BMs across all regions and quarters for which we received consumption information.

As can be seen:

- BMs with multiplicative adjustment factors always produce lower median RRMSEs than those with additive adjustment factors but none of the candidate BMs achieves a median RRMSE below 0.12 across any of the quarters within or over the entire three-year analysis. The median ARE scores within the 'good' range in all cases but tend to show a negative bias
- BMs with pre- and post-adjustment periods always provide lower median RRMSEs than those with only pre-adjustment periods; the difference in their respective median RRMSEs being about 0.01 better for the pre/post options
- A 20% adjustment factor was better in all quarters except Quarter 1 for the pre-only adjustment factor BMs, where the 40% factor produced marginally better median RRMSEs.

27 As noted in section 2.4, due to the size of this segment and the processing time required for each NMI, the analyses were undertaken using a 10% sample of all NMIs within this segment, by NEM region.

- A 40% adjustment factor provided lower median RRMSEs in the pre/post adjustment BMs in all quarters except quarter 3.

Table 4: Comparison of BMs for all years and regions, by quarter – Medium-Sized Business Customers, 3:30-8:00pm event window

Regions	Years	Qtr	Outcomes	Comments on median RRMSE outcome	Best = Multiplicative	
					Pre	Pre&Post
All	All	All	**RRMSE** medians: Max 0.151, Min 0.130. **ARE** medians: -0.002 to +0.001. **RER** median values are not significant. RERT equivalent 90th percentile: 0.629	Pre only: Multiplicative RRMSE better by ≈0.017, adjustment percentage marginal difference. Pre and post RRMSE better by 0.006 to 0.015 with multiplicative 40% best	0.139 20% 0.0	0.130 40% 0.0
All	All	1	**RRMSE** medians Max 0.159, Min 0.132. **ARE** medians: -0.004 to 0.0. **RER** median values are not significant. RERT equivalent 90th percentile 0.676	Pre only: Multiplicative RRMSE better by 0.007, Adjustment percentage of 40% marginally better. Pre and post RRMSE better by 0.07 to 0.020 with multiplicative 40% best.	0.142 40% -0.003	0.132 40% -0.001
All	All	2	**RRMSE** medians Max 0.144, Min 0.126. **ARE** medians: -0.006 to -0.001. **RER** median values are not significant. RERT 90th percentile 0.591	Pre only: Multiplicative better by 0.007 with lower adjustments 0.002 better. Pre and post RRMSE around the same and better by 0.0010 to 0.010.	0.135 20% -0.004	0.126 40% -0.001
All	All	3	**RRMSE** medians Max .136, Min .120. **ARE** medians: 0.0 to +0.002. **RER** median values are not significant. RERT 90th percentile 0.556	Pre only: Multiplicative BMs better by ≈0.005. Adjustments not significant. Pre and post RRMSE around the same ≈0.120*.	0.129 20% +0.001	0.120* 20% +0.002
All	All	4	**RRMSE** medians Max .168, Min .142. **ARE** medians: 0.0 to +0.004. **RER** median values are not significant. RERT 90th percentile 0.704	Pre only RRMSE for multiplicative better by ≈0.015. Adjustments not significant. Pre and post multiplicative with 40% cap best and better by 0.010.	0.152 20% +0.003	0.142 40% +0.003

Table 5 provides similar information for the 8 candidate BMs across all regions for all of 2019 and each quarter within it. It shows that:

- BMs with multiplicative adjustment factors produce lower median RRMSEs than those with additive adjustment factors for the year as a whole and in each quarter, but in no case is the median RRMSE lower than 0.12. Median ARE scores well within the 'good' range for all BMs in all timeframes except for an excursion just outside the range (to 0.006 for one candidate BM in the Q2
- BMs with pre- and post-adjustment periods always provide lower median RRMSEs than those with pre-only periods, and as in the three-year timeframe, the pre/post adjustment BMs tend to have median RRMSE scores that are lower by about 0.01
- A 20% adjustment factor was seen to produce lower median RRMSEs across the year as a whole and for Quarters 2, 3 & 4 for the pre-only adjustment factor BMs; the 40% factor produced lower median RRMSEs in Quarter 1
- A 40% adjustment factor provided lower median RRMSEs in all quarters for the pre/post adjustment BMs.

Table 5: Comparison of BMs for all regions for 2019 by quarter - Medium-Sized Business Customers, 3:30-8:00pm event window

Regions	Years	Qtr	Outcomes	Comments on median RRMSE outcomes	Best = Multiplicative	
					Pre	Pre&Post
All	2019	All	**RRMSE** medians Max 0.153, Min 0.131. **ARE** medians: -0.003 to +0.001. **RER** median values are marginally better for multiplicative. RERT 90th percentile 0.618.	Pre only: Multiplicative is 0.011 better with a marginal improvement for the 20% adjustment amount. Pre and Post is better by 0.006 to 0.014 with a small improvement for the higher adjustment amounts.	0.141 20% -0.001	0.131 40% 0.0
All	2019	1	**RRMSE** medians Max 0.130, Min 0.156. **ARE** medians: -0.005 to -0.010. **RER** median values are better for multiplicative. RERT 90th percentile 0.625.	Pre only: Multiplicative is ≈.016 better with a marginal improvement for the increased adjustment amount. Pre and Post is better by .009 to .014 with improvement for the higher adjustment amounts.	0.139 40% -0.004	0.130 40% -0.002
All	2019	2	**RRMSE** medians Max .145, Min 0.127. **ARE** medians: -0.006 to -0.001. **RER** median values are not significant. RERT 90th percentile 0.597.	Pre only: Multiplicative is .007 better with a slight improvement the 20% adjustment amount. Pre and Post: Multiplicative is better by .011 to .014 with marginal improvement for the higher adjustment amounts.	0.136 20% -0.004	0.127 40% -0.001
All	2019	3	**RRMSE** medians Max 0.137, Min 0.121. **ARE** medians: 0.0 to +0.002. **RER** median values are not significant. RERT 90th percentile 0.543.	Pre only: Multiplicative is ≈.006 better with marginal improvement for the 20% adjustment. Pre and Post RRMSE medians are the same* for all methods and adjustment amounts.	0.130 20% +0.001	0.121* All equal +0.002
All	2019	4	**RRMSE** medians Max 0.172, Min 0.146. **ARE** medians: -0.001 to +0.002. **RER** median values are better for multiplicative. RERT 90th percentile 0.706.	Pre only: Multiplicative is 0.013 better with a no improvement for the increased adjustment amount. Pre and Post is better by 0.10 to 0.015 with multiplicative 40% significantly better.	0.158 20% 0.0	0.146 40% +0.002

3.1.2. Assessment of eligibility of Medium-Sized Business Customers at various accuracy/frequency thresholds

Table 6 provides the results of this analysis for all of 2019 across all regions. As can be seen:

- The 0.1/100% threshold would result in between 3.7% and 4.0% of the NMIs in Segment E being eligible to participate in the WDRM under the pre-only adjustment BMs and from 5.1% to 5.8% under the pre/post BMs
- The 0.2/100% threshold would significantly increase eligibility in both types of BMs – by a factor in the range of 4 to 5 for the pre-only BMs and about 5 in the case of the pre/post BMs, but would double the level of expected error entailed
- The 0.1/90% threshold would result in eligibility levels over 4 times higher than the 0.1/100% threshold for the pre-only BMs and about 3 times higher for the pre/post BMs.

Table 6: Eligible percentage for different required frequencies of producing a threshold RRMSE - Medium-Sized Business Customers, 3:30-8:00pm event window, all quarters of 2019, NEM-wide

RRMSE (BM%)	Pre only adjustments				Pre and Post adjustments			
	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%
0.1 100%	3.7	**4.0**	3.7	3.9	5.4	**5.8**	5.1	5.4
0.2 100%	18.4	17.3	**20.0**	18.9	26.0	**26.6**	26.0	26.1
0.1 90%	12.4	**12.6**	12.5	12.5	15.7	16.7	15.8	**16.8**

Table 7 provides similar information for the 0.1/100% threshold for all quarters of 2019 for each of the regions. It shows that:

- The levels of eligibility for Segment E under this threshold are very low in all NEM regions for all of the candidate BMs; in no case does eligibility reach even 5.5%.
- Although the pre/post adjustment BMs would allow a greater proportion of NMIs in three of the regions (New South Wales, Queensland and Victoria), the increase in eligibility is less than 2% in all cases. In the other two regions (South Australia and Tasmania), there is no or very little difference between the pre/post and pre-only adjustment BMs
- There is no clear advantage in terms of eligibility for any particular BM within either the pre-only or pre/post BMs.

Table 7: Eligible percentage where the NMI is required to provide an RRMSE of 0.1 or less with 100% frequency - Medium-Sized Business Customers, 3:30-8:00pm event window, all quarters of 2019 by region

Region	Pre only adjustments				Pre and Post adjustments			
	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%
NSW	3.7	4.0	3.7	**3.9**	5.4	**5.8**	5.1	5.4
QLD	2.1	**2.2**	2.1	2.1	3.8	**4.0**	3.7	3.8
SA	1.5	1.5	1.5	1.5	1.5	1.5	**1.5**	**1.5**
TAS	2.7	2.7	2.7	2.7	2.7	2.7	**3.4**	**3.4**
VIC	**3.4**	3.4	**3.4**	**3.4**	4.6	**4.6**	4.3	4.2

Table 8 provides similar information for the 0.2/100% threshold. It shows that:

- It would significantly increase eligibility:
  - Under the pre-only BMs eligibility would increase by factors ranging from over 4 in the case of New South Wales to close to 10 for South Australia
  - Under the pre/post BMs the increases would be even greater, ranging from a factor of about 5 in New South Wales to more than 10 in South Australia.
- However, the use of the 0.2 RRMSE threshold would essentially double the potential expected level of error in the administration of the WDRM as compared to the 0.1 level.

Table 8: Eligible percentage where the NMI is required to provide an RRMSE of 0.2 or less with 100% frequency - Medium-Sized Business Customers, 3:30-8:00pm event window, all quarters of 2019 by region

Region	Pre only adjustments				Pre and Post adjustments			
	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%
NSW	17.6	16.5	**17.7**	17.1	**25.8**	25.7	24.4	23.9
QLD	18.4	17.3	**20.0**	18.9	26.0	**26.6**	26.0	26.0
SA	**14.1**	13.8	13.8	12.9	**18.1**	17.3	17.7	17.4
TAS	12.1	10.7	**12.8**	11.4	**17.5**	**17.5**	16.1	**17.5**
VIC	18.4	17.1	**18.7**	17.8	**24.4**	24.3	23.5	23.7

Table 9 shows the same information for the 0.1/90% threshold, and shows that it would:

- Increase the proportion of eligible NMIs in each region significantly as compared to 0.1/100% level
- However, eligible proportions in South Australia and Tasmania would remain below 10%, ranging from about 6% to no more than just over 9% under any of the BMs; but contrast, in the other regions the eligible proportion would range from about 10.5% to just under 16% depending on the region and BM
- Eligible proportions would be consistently higher under the pre/post BMs than the pre-only versions.

Table 9: Eligible percentage where the NMI is required to provide an RRMSE of 0.1 or less with 90% frequency - Medium-Sized Business Customers, 3:30-8:00pm event window, all quarters of 2019 by region

Region	Pre only adjustments				Pre and Post adjustments			
	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%
NSW	12.4	**12.6**	12.5	12.5	15.7	**15.9**	15.2	15.6
QLD	10.5	10.4	**12.0**	11.7	15.7	16.7	15.8	**16.8**
SA	6.5	7.1	**7.3**	**7.3**	**9.1**	**9.1**	8.2	8.5
TAS	6.0	6.0	**6.7**	**6.7**	**8.1**	**8.1**	7.4	7.4
VIC	10.7	10.9	11.1	**11.2**	15.1	**15.2**	14.2	14.1

3.2. Short early evening event window (5:30-7:00pm)

3.2.1. Assessment of accuracy, bias and variability of BMs for Medium-Sized Business Customers

Table 10 below provides the results of this analysis for all 8 candidate BMs across all regions and quarters for which we received consumption information.

As can be seen:

- None of the BMs exhibit median RRMSEs within the 'good' level, though all are well within the 'acceptable' range in all regions and years

- The BMs with multiplicative adjustment factors always produce median RRMSEs that are lower or equal to those with additive adjustment factors; median ARE scores are well within the 'good' range in almost all cases, with only two cases at or just above the 'good' threshold
- BMs with pre- and post-adjustment periods always provide lower median RRMSEs than those with only pre-adjustment periods
- A 20% adjustment factor was better in all quarters except Quarter 1 for the pre-only adjustment factor BMs, where the 40% factor produced marginally lower median RRMSEs
- A 40% adjustment factor provided lower median RRMSEs in the pre/post adjustment BMs in all quarters.

Table 10: Comparison of BMs for all years and regions, by quarter – Medium-Sized Business Customers, 5:30-7:00pm event window

Regions	Years	Qtr	Outcomes	Comments on median RRMSE outcome	Best = Multiplicative	
					Pre	Pre&Post
All	All	All	**RRMSE** medians: Max 0.133, Min 0.109. **ARE** medians: -0.002 to +0.002. **RER** median values are not significant. RERT equivalent 90th percentile: 0.713	Pre only: Multiplicative RRMSE 20% better by 0.010, adjustment percentage of marginal difference. Pre and post RRMSE better by ≈ 0.015 with multiplicative 40% better	0.123 20% +0.001	0.109 40% +0.002
All	All	1	**RRMSE** medians Max 0.141, Min 0.111. **ARE** medians: -0.003 to +0.002. **RER** median values are not significant. RERT equivalent 90th percentile 0.806	Pre only: Multiplicative RRMSE better by 0.015, Adjustment percentage of 20% marginally better. Pre and post RRMSE better by 0.015 to 0.019 with multiplicative and 40% caps better	0.125 40% +0.001	0.111 40% +0.001
All	All	2	**RRMSE** medians Max 0.129, Min 0.105. **ARE** medians: +0.006 to -0.001. **RER** median values are not significant. RERT 90th percentile 0.653	Pre only multiplicative RRMSE .007 better, with 20% adjustment best. All pre and post RRMSE around the same with multiplicative 40% better.	0.120 20% -0.004	0.105 40% -0.001
All	All	3	**RRMSE** medians Max .122, Min .101. **ARE** medians: +0.002 to +0.004. **RER** median values are not significant. RERT 90th percentile 0.591	Pre only: multiplicative .006 better with 20% adjustment best. Pre and post: three RRMSE the same* at 0.101, with additive 20% .002 worse.	0.114 20% +0.003	0.101* 40% +0.002
All	All	4	**RRMSE** medians Max .150, Min .119. **ARE** medians: +0.001 to +0.005. **RER** median values are not significant. RERT 90th percentile 0.819.	Pre only: RRMSE for multiplicative better for pre only adjustment by 0.015 with the 20% adjustment better again. Pre and post: Better by 0.016 to 0.021. multiplicative with 40% cap best.	0.132 20% +0.005	0.119 40% +0.004

Table 11 provides similar information for the 8 candidate BMs across all regions for all of 2019 and each quarter within it. It shows that:

- BMs with multiplicative adjustment factors produce lower median RRMSEs than those with additive adjustment factors in all but one case, Q3, in which an additive adjustment factor produced equally good median RRMSEs in the pre/post BMs. In all other cases, the 40% adjustment provides lower median RRMSEs in the pre/post BMs, while the 20% adjustment factor does so in the pre-only BMs.
- All median ARE scores except one are within the 'good' range; the exception of +0.008 occurs in Q2 but is still within the 'acceptable' range
- BMs with pre- and post-adjustment periods always provide lower median RRMSEs than those with only pre-adjustment periods

Table 11: Comparison of BMs for all regions for 2019 by quarter – Medium-Sized Business Customers, 5:30-7:00pm event window

Regions	Years	Qtr	Outcomes	Comments on median RRMSE outcomes	Best = Multiplicative Pre	Pre&Post
All	2019	All	**RRMSE** medians Max 0.136, Min 0.109. **ARE** medians: -0.002 to +0.002. **RER** median values are not significant. RERT 90th percentile 0.693.	Pre only: Multiplicative is ≈ 0.010 better with 20% adjustment 0.002 better. Pre and Post is better by ≈ 0.016 with multiplicative 40% best.	0.124 20% 0.0	0.109 40% +0.001
All	2019	1	**RRMSE** medians Max 0.139, Min 0.110. **ARE** medians: -0.004 to -0.001. **RER** median values are not significant. RERT 90th percentile 0.737.	Pre only multiplicative is .012 better with no difference for the adjustment amount. Pre and Post is better by .014 to .018 with multiplicative 40% best.	0.124 20% 0.0	0.110 40% +0.001
All	2019	2	**RRMSE** medians Max .132, Min 0.106. **ARE** medians: +0.008 to -0.001. **RER** median values are not significant. RERT 90th percentile 0.660.	Pre only multiplicative is ≈0.008 better with 20% adjustment slightly better. Pre and Post is better by 0.015 to 0.020 with multiplicative 40% best.	0.121 20% -0.005	0.106 40% -0.001
All	2019	3	**RRMSE** medians Max 0.123 Min 0.101. **ARE** medians: 0.001 to 0.003. **RER** median values are not significant. RERT 90th percentile 0.582.	Pre only multiplicative with 20% adjustment is 0.005. Pre and Post is better by ≈ 0.018 with all bar additive* 20% the same at 0.101.	0.115 20% +0.003	0.101* 20% +0.003
All	2019	4	**RRMSE** medians Max 0.152, Min 0.120. **ARE** medians: 0.0 to +0.004. **RER** median values are not significant. RERT 90th percentile 0.793.	Pre only: multiplicative is ≈ 0.020 better with 20% adjustment 0.002 better. Pre and Post is better by 0.018 to 0.021 with multiplicative 40% best.	0.136 20% +0.003	0.120 40% +0.003

3.2.2. Assessment of eligibility of Medium-Sized Business Customers at various accuracy/frequency thresholds

Table 12 provides the results of the analysis of the proportion of the NMIs within this segment that would be eligible for the WDRM for all of 2019 across all regions. As can be seen:

- The 0.1/100% threshold would result in between about 5% of the NMIs in Segment E being eligible to participate in the WDRM under the pre-only adjustment BMs and about 6.5% under the pre/post BMs
- The 0.2/100% threshold would increase the proportion of eligible NMIs by a factor of approximately 3 for the pre-only BMs and a factor of about 4 for the pre/post BMs, but would double the potential total level of error entailed
- The 0.1/90% threshold would result in eligibility levels closer to the 0.2/100% threshold level than the 0.1/100% level, with the proportion of eligible NMIs under the pre-only BMs ranging from about 13% to 13.5% and from about 17.5% to 19% for the pre/post BMs. However, as noted above, this threshold would not, in ad of itself, cap the total level of potential error that could be encountered.

Table 12: Eligible percentage for different required frequencies of producing a threshold RRMSE - Medium-Sized Business Customers, 5:30-7:00pm event window, all quarters of 2019, NEM-wide

RRMSE (BM%)	Pre only adjustments				Pre and Post adjustments			
	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%
0.1 100%	**5.1**	**5.1**	5.0	4.9	6.4	6.6	6.5	**6.6**
0.2 100%	**18.7**	17.2	17.9	16.8	27.3	28.1	26.4	**27.0**
0.1 90%	**13.5**	13.4	13.0	12.9	18.0	18.6	17.7	**19.2**

Table 13 provides similar information for the 0.1/100% threshold for all quarters of 2019 for each of the regions. It shows that:

- The pre/post adjustment BMs would allow a materially greater proportion of the Segment E NMIs to be eligible to participate than the pre-only BMs in all NEM regions except South Australia, where they would be essentially the same. However, eligibility levels would be low in all regions ranging from just over 2% in South Australia to just over 5% in New South Wales for the pre-only BMs, and between just over 2% in South Australia to just over 6.6% in New South Wales for the pre/post BMs.
- There is not a material spread of the eligible proportion across the pre-only or the pre/post BMs ***within*** each of the NEM regions; the difference ***across*** the regions is larger, but still not particularly large, being a factor of about 2 in the pre-only BMs and as much as 3 in the pre/post BMs.:

Table 13: Eligible percentage where the NMI is required to provide an RRMSE of 0.1 or less with 100% frequency - Medium-Sized Business Customers, 5:30-7:00pm event window, all quarters of 2019 by region

Region	Pre only adjustments				Pre and Post adjustments			
	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%
NSW	**5.1**	**5.1**	5.0	4.9	6.4	6.6	6.5	**6.6**
QLD	**2.5**	**2.5**	2.3	2.3	4.6	**4.9**	4.6	4.8
SA	**2.3**	**2.3**	2.1	2.1	**2.1**	**2.1**	1.8	**2.06**
TAS	2.7	2.7	2.7	2.7	2.7	2.7	**2.7**	**2.7**
VIC	3.1	3.2	**3.4**	**3.4**	4.3	**4.6**	4.3	4.4

Table 14 provides similar information for the 0.2/100% threshold. It shows that:

- It would significantly increase eligibility:
  - Under the pre-only BMs eligibility would increase by factors ranging from about 3 to 4 in the case of New South Wales to almost 7 for Queensland
  - Under the pre/post BMs the increases would be materially higher – ranging from a factor of 4 to 5 for New South Wales, Queensland and Victoria to about 6 to 8 or above for Tasmania and South Australia respectively
- But, as noted above, the use of the 0.2 RRMSE would essentially double the expected level of error in the administration of the WDRM.

Table 14: Eligible percentage where the NMI is required to provide an RRMSE of 0.2 or less with 100% frequency - Medium-Sized Business Customers, 5:30-7:00pm event window, all quarters of 2019 by region

Region	Pre only adjustments				Pre and Post adjustments			
	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%
NSW	**18.7**	17.2	17.9	16.8	27.3	**28.1**	26.4	27.0
QLD	16.1	15.7	**16.8**	16.6	26.4	**26.9**	25.8	26.4
SA	**13.8**	11.8	11.2	10.6	**19.9**	19.0	18.4	18.7
TAS	10.3	8.9	**11.6**	9.6	15.8	**16.8**	15.1	16.1
VIC	**17.1**	16.1	16.5	15.8	**23.9**	23.6	22.7	23.8

Table 15 shows the same information for the 0.1/90% threshold, and shows results similar to those across the regions in that it would:

- Result in proportions of eligible NMIs being much closer to those of the 0.2/100% threshold than the 0.1/100% threshold in New South Wales, Queensland and Victoria, and closer, though less so, for South Australia and Tasmania
- Eligible proportions would range from about 6% to over 7.5% in Tasmania to about 13% to 13.5% for NSW in the pre-only BMs and from around 9% for Tasmania to about 18% for NSW under the pre/post BMs.

Table 15: Eligible percentage where the NMI is required to provide an RRMSE of 0.1 or less with 90% frequency - Medium-Sized Business Customers, 5:30-7:00pm event window, all quarters of 2019 by region

Region	Pre only adjustments				Pre and Post adjustments			
	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%
NSW	**13.5**	13.4	13.0	12.9	18.0	**18.6**	17.7	**17.8**
QLD	11.6	11.5	**12.6**	12.3	17.5	18.5	17.5	**19.2**
SA	**8.2**	7.9	7.9	7.9	10.5	**11.4**	10.2	11.1
TAS	6.0	6.7	6.7	**7.4**	8.7	8.7	8.7	**9.4**
VIC	10.8	11.0	11.4	**11.6**	15.9	**16.3**	14.8	15.3

3.3. Morning event window (7:00-9:00am)

3.3.1. Assessment of accuracy, bias and variability of BMs for Medium-Sized Business Customers

Table 16 below provides the results of this analysis for all 8 candidate BMs across all regions and quarters for which we received consumption information.

As can be seen:

- BMs with pre- and post-adjustment periods always provide lower median RRMSEs than those with only pre-adjustment periods, but none of the candidate BMs exhibit median RRMSE scores that fall within the 'good' range, though all fall well within the 'acceptable' range

- Pre-only RRMSEs, across all regions and all years, had a median of 0.0123, whereas for the pre/post BMs this value was 0.0115
- Median AREs were in the 'good' range for both types of BMs
- BMs with additive adjustment factors were seen to produce better median RRMSEs for the pre-only BMs, but the improvement in the median RRMSE was never more than 0.015; the difference in the median RRMSEs of the additive and multiplicative adjustments in the pre/post BMs was even smaller
- A 20% adjustment factor produced lower median RRMSEs for the pre-only BMs overall, but a 40% factor provided lower median RRMSEs in quarters 1 and 2
- A 40% adjustment factor provided lower median RRMSEs in the pre/post adjustment BMs in all quarters except quarter 3.

Table 16: Comparison of BMs for all years and regions, by quarter – Medium-Sized Business Customers, 7:00-9:00am event window

Regions	Years	Qtr	Outcomes	Comments on median RRMSE outcomes	Best = Additive	
					Pre	Pre&Post
All	All	All	**RRMSE** medians: Max 0.136, Min 0.115. **ARE** medians: -0.001 to +0.002. **RER** median values are not significant. RERT equivalent 90th percentile: 0.501	Pre only additive* RRMSE better by 0.010 to 0.013. No difference for adjustment percentage. Pre and post RRMSE better by 0.005 with the alternatives equal but the higher adjustments marginally better.	0.123 20% 0.002	0.115 40% -0.001
All	All	1	**RRMSE** medians Max 0.136, Min 0.112. **ARE** medians: -0.002 to 0.003. **RER** median values are not significant. RERT equivalent 90th percentile 0.504	Pre only additive RRMSE better by 0.007, Adjustment percentage of 40% better by 0.003. Pre and post RRMSE better by 0.013 to 0.016 with multiplicative and additive around the same. 40% caps better	0.126 40% 0.003	0.112 40% -0.002
All	All	2	**RRMSE** medians Max 0.135, Min 0.118. **ARE** medians: -0.003 to -0.001. **RER** median values are not significant. RERT 90th percentile 0.495	Pre only additive RRMSE 0.012 better, with 40% adjustments marginally better. All pre and post RRMSE are the same but multiplicative have a marginally better median ARE*.	0.120 40% -0.001	0.118 20% -0.002*
All	All	3	**RRMSE** medians Max .127, Min .110. **ARE** medians: +0.002 to +0.004. **RER** median values are not significant. RERT 90th percentile 0.454.	Pre only additive BMs better by 0.015. No difference for adjustments. All pre and post RRMSE around the same ≈0.011 with additive 40% best.	0.112 20% +0.004	0.110 40% +0.002
All	All	4	**RRMSE** medians Max .147, Min .119. **ARE** medians: 0.0 to +0.001. **RER** median values indicated pre- and post- multiplicative marginally better. RERT 90th percentile 0.552.	Pre only additive RRMSE better by ≈0.011 with 20% adjustment best. Pre and post better by 0.011 to 0.024 with additive and multiplicative the same. 40% cap is best.	0.132 20% +0.001	0.119 40% +0.001

Table 17 provides similar information for the 8 candidate BMs across all regions for all of 2019 and each quarter within it. It shows that:

- Although the pre/post BMs always provided lower median RRMSEs than the pre-only BMs, none of the candidate BMs exhibited median RRMSE scores that fell within the 'good' range, though all fell well within the 'acceptable' range
- All median ARE scores were within the 'good' range, though in two cases they were just at the edge (-0.005 in one case and +0.005 in the other)

- BMs with additive adjustment factors were seen to produce better median RRMSEs for the pre-only BMs, but the improvement in the median RRMSE was never more than 0.013; the difference between the median RRMSEs of the additive and multiplicative adjustments in the pre/post BMs was even smaller
- A 40% adjustment factor provided lower median RRMSEs across the year and in Quarters 1, 2 and 3 for the pre-only BMs; the 40% adjustment factor provided lower median RRMSEs in all quarters for the pre/post BMs.

Table 17: Comparison of BMs for all regions for 2019 by quarter – Medium-Sized Business Customers, 7:00-9:00am event window

Regions	Years	Qtr	Outcomes	Comments on median RRMSE outcomes	Best = Additive	
					Pre	Pre&Post
All	2019	All	**RRMSE** medians Max 0.137, Min 0.116. **ARE** medians: -0.001 to +0.001. **RER** median values are not significant. RERT 90th percentile 0.498.	Pre only: additive is ≈ 0.010 better with no material change for adjustment %. Pre and Post is better by 0.007 to 0.024 with additive and multiplicative equal* and 40% adjustment marginally better.	0.123 40% +0.002	0.116 40% -0.001
All	2019	1	**RRMSE** medians Max 0.133, Min 0.110. **ARE** medians: -0.003 to -0.001. **RER** median values are not significant. RERT 90th percentile 0.479.	Pre only additive is 0.007 better with a marginal improvement for the increased adjustment amount. Pre and Post is better by .012 to .023 with multiplicative* 40% marginally best.	0.124 40% +0.001	0.110* 40% -0.002
All	2019	2	**RRMSE** medians Max .136, Min 0.118. **ARE** medians: -0.001 to -0.005. **RER** median values are not significant. RERT 90th percentile 0.501.	Pre only additive is ≈0.011 better with a marginal improvement for 40% adjustment. Pre and Post is better by .03 with additive 40% adjustment best.	0.121 40% -0.001	0.118 40% -0.004
All	2019	3	**RRMSE** medians Max 0.126, Min 0.110. **ARE** medians: +0.002 to 0.005. **RER** median values are not significant. RERT 90th percentile 0.449.	Pre only additive is 0.013 better with no difference for adjustment percentage. Pre and Post is marginally better for additive and by ≈ 0.014 for multiplicative. Additive 40% is marginally better than the others.	0.111 40% 0.004	0.110 40% 0.002
All	2019	4	**RRMSE** medians Max 0.151, Min 0.124. **ARE** medians: 0.0 to +0.002. **RER** median values are not significant. RERT 90th percentile 0.564.	Pre only: additive is ≈ 0.009 better with a no improvement for the increased adjustment amount. Pre and Post is better by 0.011 to 0.022. Additive and multiplicative 40% are equally best*.	0.137 20% +0.001	0.124* 40% 0.0

3.3.2. Assessment of eligibility of Medium-Sized Business Customers at various accuracy/frequency thresholds

The tables in this section provide the following information:

Table 18 provides this information for all of 2019 across all regions. As can be seen:

- The 0.1/100% threshold would result in between 3.1% and about 3.8% of the NMIs in Segment E being eligible to participate in the WDRM under the pre-only adjustment BMs and about 4.0% to 4.5% under the pre/post BMs
- The 0.2/100% threshold would increase that proportion by a factor of about 4 for the pre-only BMs and a factor of about 5 for the pre/post BMs, but would double the total potential error entailed in the use of these BMs
- The 0.1/90% threshold would result in eligibility levels closer to the 0.2/100% threshold level.

Table 18: Eligible percentage for different required frequencies of producing a threshold RRMSE - Medium-Sized Business Customers, 5:30-7:00pm event window, all quarters of 2019, NEM-wide

RRMSE (BM%)	Pre only adjustments				Pre and Post adjustments			
	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%
0.1 100%	**3.8**	3.7	3.2	3.2	4.3	**4.5**	4.0	**4.1**
0.2 100%	**17.9**	17.3	14.4	13.5	**22.0**	21.7	21.4	21.4
0.1 90%	**12.2**	12.3	9.9	9.9	16.0	**16.4**	15.5	16.2

Table 19 provides similar information for the 0.1/100% threshold for all quarters of 2019 for each of the regions. It shows that:

- The pre/post adjustment BMs would all but two cases allow a greater proportion of NMIs to be eligible to participate in the WDRM than the pre-only BMs in every NEM region. However, eligibility would be very low in all cases, never reaching 4% in the pre-only BMs and only approaching 4.5% in one case under a pre/post BM
- There is only a small spread of the eligible proportion across the pre-only or the pre/post BMs ***within*** each of the NEM regions (never even as much as 1%); the difference between the pre-only and pre/post BMs ***across*** the regions is also relatively modest – never more that about 2.5% in the pre-only BMs and never more that 3.5% in the pre/post BMs.

Table 19: Eligible percentage where the NMI is required to provide an RRMSE of 0.1 or less with 100% frequency - Medium-Sized Business Customers, 7:00-9:00am event window, all quarters of 2019 by region

Region	Pre only adjustments				Pre and Post adjustments			
	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%
NSW	**3.8**	3.7	3.2	3.20	4.3	**4.5**	4.0	4.1
QLD	1.6	1.6	1.6	**1.7**	2.6	**2.7**	2.6	**2.7**
SA	0.9	0.9	0.9	0.9	**1.5**	1.2	1.2	1.2
TAS	1.8	1.8	1.8	1.8	2.6	2.6	2.6	2.6
VIC	**3.0**	2.8	2.1	2.0	**2.9**	**2.9**	2.5	2.5

Table 20 provides similar information for the 0.2/100% threshold. It shows that:

- It would significantly increase eligibility:
  - Under the pre-only BMs eligibility would increase significantly, ranging from a factor of 4 to 5 in the case of New South Wales to factors of more than 7 for South Australia and 10 for Queensland
  - Similar and even higher multiples would be produced in the pre/post BMs
- But, as noted above, the use of the 0.2 RRMSE level would essentially double the total potential error in the WDRM.

Table 20: Eligible percentage where the NMI is required to provide an RRMSE of 0.2 or less with 100% frequency - Medium-Sized Business Customers, 7:00-9:00am event window, all quarters of 2019 by region

Region	Pre only adjustments				Pre and Post adjustments			
	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%
NSW	**17.9**	17.3	14.4	13.3	**21.8**	21.7	21.4	21.4
QLD	**16.5**	15.8	14.1	13.5	**22.0**	21.3	21.2	19.8
SA	**11.1**	10.2	8.2	6.8	11.1	**11.7**	9.9	10.8
TAS	**12.2**	11.4	7.1	7.0	14.3	16.1	15.8	**16.7**
VIC	**15.8**	15.0	12.6	12.2	18.5	**18.7**	18.0	18.3

Table 21 shows the same information for the 0.1/90% threshold, and shows results similar to those across the regions in that it would:

- Result in proportions of eligible NMIs being much closer to those of the 0.2/100% threshold than the 0.1/100% threshold in New South Wales, and about midway between the eligibility proportions achieved in the other regions under the two other thresholds
- Eligible proportions would range from about 3.5% to 6.1% in Tasmania to about 10% to over 12% for NSW in the pre-only BMs and from around 5.5% for Tasmania to about 15.5% to 16.5% for NSW under the pre/post BMs.

Table 21: Eligible percentage where the NMI is required to provide an RRMSE of 0.1 or less with 90% frequency - Medium-Sized Business Customers, 7:00-9:00am event window, all quarters of 2019 by region

Region	Pre only adjustments				Pre and Post adjustments			
	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%
NSW	12.2	**12.3**	9.9	9.9	16.0	**16.4**	15.5	16.2
QLD	8.6	**8.7**	7.0	7.1	**14.5**	13.9	14.3	13.8
SA	7.0	**7.3**	5.6	5.6	**5.9**	**5.9**	5.3	5.6
TAS	**6.1**	5.3	3.6	3.5	8.9	**9.8**	8.9	8.9
VIC	**10.2**	9.8	7.4	7.6	**11.1**	**11.1**	10.7	10.7

4. Results for Large Commercial and Industrial Customers

This section of the report provides the results for Large Commercial and Industrial Customers – those customers with annual electricity consumption ranging from 750 MWh to 100GWh. This segment contains the majority of the large commercial and industrial customers in the NEM.

The results of the two levels of analysis are discussed for each of the event windows identified as containing high prices. The extended afternoon/evening event window (which runs from 3:30pm to 8:00pm) is discussed first as it contains the largest proportion of high-price trading intervals annually. Subsequent sections provide parallel discussions of the short early evening event window (which runs from 5:30pm to 7:00pm) and the morning event window (which runs from 7:00am to 9:00am).

Within each event window the results of the analysis are provided as follows:

- First the results of the assessment of the accuracy, bias and variability of each of the various BMs when applied to this customer segment are discussed
- Then the results are discussed of the analysis of the percentage of the NMIs within this segment that would be expected to be eligible to participate in the WDRM under the three different combinations of threshold RRMSE and the frequency with which the NMI met that threshold in the preceding calendar quarter. It should be noted that the analysis of eligibility is discussed only for the afternoon/evening price event window in order to reduce the length of this section of the report. Appendix D provides links to spreadsheets that provide the eligihle percentage of each customer segments in each NEM region for each of the three eligibility thresholds analysed.

Details of the contents of the tables used in the two analyses are provided in section 2.6.4 above. It should be noted that the

4.1. Afternoon/evening event window (3:30-8:00pm)

4.1.1. Assessment of accuracy, bias and variability of BMs for Large Commercial and Industrial Customers

Table 22 below provides the results of this analysis for all 8 candidate BMs across all regions and quarters for which we received consumption information.

As can be seen:

- BMs with multiplicative adjustment factors always produce lower median RRMSEs than those with additive adjustment factors, and have median ARE scores well within the ‘good’ range
- BMs with pre- and post-adjustment periods always provide lower median RRMSEs than those with only pre-adjustment periods
- A 20% adjustment factor provided lower median RRMSEs in all quarters except Quarter 1 for the pre-only adjustment factor BMs, where the 40% factor produced marginally better median RRMSEs.
- A 40% adjustment factor provided lower median RRMSEs in the pre/post adjustment BMs in all quarters except Quarter 3.

Table 22: Comparison of BMs for all years and regions, by quarter – Large Commercial and Industrial Customers, 3:30-8:00pm event window

Regions	Years	Qtr	Outcomes	Comments on median RRMSE outcome	Best = Multiplicative	
					Pre	Pre&Post
All	All	All	**RRMSE** medians: Max 0.09, Min 0.076. **ARE** medians: -0.001 to +0.001. **RER** median values are not significant. RERT equivalent 90th percentile: 0.425	Pre only: Multiplicative RRMSE better by 0.005, adjustment percentage no real difference. Pre and post RRMSE better by 0.006 to 0.008 with multiplicative marginally better	0.085 20% -0.001	0.077 40% +0.001
All	All	1	**RRMSE** medians Max 0.097, Min 0.079. **ARE** medians: -0.003 to +0.001. **RER** median values are not significant. RERT equivalent 90th percentile 0.458	Pre only: Multiplicative RRMSE better by 0.007, Adjustment percentage of 40% marginally better. Pre and post RRMSE better by 0.012 to 0.006 with multiplicative marginally better	0.090 40% -0.002	0.079 40% -0.001
All	All	2	**RRMSE** medians Max 0.084, Min 0.072. **ARE** medians: -0.004 to 0.0. **RER** median values are not significant. RERT 90th percentile 0.398	All pre only BMs around the same RRMSE. All pre and post RRMSE around the same.	0.082 20% -0.002	0.072 40% 0.0
All	All	3	**RRMSE** medians Max .081, Min .070. **ARE** medians: +0.001 to +0.002. **RER** median values are not significant. RERT 90th percentile 0.385	All pre only BMs around the same RRMSE ≈ 0.08. All pre and post RRMSE around the same ≈ 0.07.	0.079 20% +0.001	0.070 20% +0.002
All	All	4	**RRMSE** medians Max .099, Min .083. **ARE** medians: +0.001 to +0.003. **RER** median values are not significant. RERT 90th percentile 0.467	Pre only RRMSE for multiplicative better for pre only adjustment – 0.092 compared to 0.099 with the adjustment % not varying. Pre and post multiplicative with 40% cap best.	0.092 20% +0.001	0.083 40% +0.003

Table 23 provides similar information for the 8 candidate BMs across all regions for all of 2019 and each quarter within it. It shows that:

- BMs with multiplicative adjustment factors produce lower median RRMSEs than those with additive adjustment factors for the year as a whole and in each quarter, and have median ARE scores well within the 'good' range
- BMs with pre- and post-adjustment periods always provide lower median RRMSEs than those with only pre-adjustment periods
- A 40% adjustment factor provided lower median RRMSEs in Quarters 1 and 4 and across the year as a whole for the pre-only adjustment factor BMs, whereas the 40% factor produced better median RRMSEs in Quarters 2 and 3
- A 40% adjustment factor provided lower median RRMSEs in all quarters for the pre/post adjustment BMs.

Table 23: Comparison of BMs for all regions for 2019 by quarter - Large Commercial and Industrial Customers, 3:30-8:00pm event window

Regions	Years	Qtr	Outcomes	Comments on median RRMSE outcomes	Best = Multiplicative Pre	Best = Multiplicative Pre&Post
All	2019	All	**RRMSE** medians Max 0.092, Min 0.077. **ARE** medians: -0.001 to +0.001. **RER** median values are not significant. RERT 90th percentile 0.417.	Pre only: Multiplicative is 0.004 better with a marginal improvement for the increased adjustment amount. Pre and Post is better by 0.004 to 0.011 with a small improvement for the higher adjustment amounts.	0.087 40% -0.001	0.077 40% +0.001
All	2019	1	**RRMSE** medians Max 0.097, Min 0.079. **ARE** medians: -0.003 to 0.0. **RER** median values are marginally better for multiplicative. RERT 90th percentile 0.451.	Pre only: Multiplicative is .008 better with a marginal improvement for the increased adjustment amount. Pre and Post is better by .009 to .011 with improvement for the higher adjustment amounts.	0.088 40% -0.002	0.079 40% -0.001
All	2019	2	**RRMSE** medians Max .087, Min 0.073. **ARE** medians: -0.004 to 0.00. **RER** median values are not significant. RERT 90th percentile 0.395.	Pre only: Multiplicative is .003 better with no change for the increased adjustment amount. Pre and Post: Multiplicative is better by .011 to .012 with marginal improvement for the higher adjustment amounts.	0.084 20% -0.003	0.073 40% 0.0
All	2019	3	**RRMSE** medians Max 0.082, Min 0.071. **ARE** medians: 0.002. **RER** median values are not significant. RERT 90th percentile 0.374.	Pre only: Multiplicative is .001 better with no change for the increased adjustment amount. Pre and Post is better by 0.01 to 0.011 with marginal improvement for the higher adjustment amounts.	0.081 20% 0.002	0.071 40% 0.002
All	2019	4	**RRMSE** medians Max 0.103, Min 0.086. **ARE** medians: 0.0 to +0.002. **RER** median values are better for multiplicative. RERT 90th percentile 0.450.	Pre only: Multiplicative is 0.006 better with a marginal improvement for the increased adjustment amount for pre only. Pre and Post is better by 0.01 to 0.012 with some improvement for the higher adjustment amounts.	0.096 40% +0.001	0.086 40% +0.002

4.1.2. Assessment of eligibility of Large Commercial and Industrial Customers at various accuracy/frequency thresholds

The tables in this section provide the following information:

- The first column identifies the RRMSE accuracy and frequency threshold for which results are described in the corresponding row. These thresholds were chosen for the following reasons:
  - The 0.1 RRMSE at 100% frequency threshold requires that the NMI achieves a 'good'[28] level of accuracy (±10%) for the baseline calculated for it over the timeframe in question
  - The 0.2 RRMSE at 100% frequency threshold requires the same level of frequency but would allow it to be at the ±20% level of accuracy, which was deemed 'acceptable' in the ARENA in its SN DR RERT Trial program
  - The 0.1 RRMSE at 90% frequency threshold represents a means for increasing eligibility levels while ensuring that level is achieved 90% of the time[29].

28 As defined in a number of studies concerning baselines including work undertaken by KEMA for AEMO in assessing BLMs for the RERT and by OGW in assessing BLMs for ARENA in its SN DR RERT Trial program.

29 It should be noted that the level of error in the remaining 10% of the baselines would not necessarily be known in this case. Ways in which this error could be bounded will be considered in Phase 2 of this study.

- The remaining columns show the proportion of the NMIs within Segment F that would meet those thresholds within the region(s) and timeframe being assessed.

Table 24 provides this information for all of 2019 across all regions. As can be seen:

- The 0.1/100% threshold would result in between 11.5% and 12.0% of the NMIs in Segment F being eligible to participate in the WDRM under the pre-only adjustment BMs and about 16.5% under the pre/post BMs
- The 0.2/100% threshold would approximately triple eligibility in both types of BMs, but would double the level of expected error entailed
- The 0.1/90% threshold would result in eligibility levels closer to the 0.2/100% threshold level.

Table 24: Eligible percentage for different required frequencies of producing a threshold RRMSE – Large Commercial and Industrial Customers, 3:30-8:00pm event window, all quarters of 2019, NEM-wide

	Pre only adjustments				Pre and Post adjustments			
RRMSE (BM%)	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%
0.1 100%	11.7	11.7	11.9	**11.9**	16.5	16.6	16.4	**16.8**
0.2 100%	35.8	33.8	**37.5**	35.4	45.1	46.1	45.0	**46.3**
0.1 90%	29.5	29.7	31.1	**31.5**	39.8	**40.6**	39.4	40.0

Table 25 provides similar information for the 0.1/100% threshold for all quarters of 2019 for each of the regions. It shows that:

- The pre/post adjustment BMs would allow a materially greater proportion of NMIs to be eligible to participate in the WDRM than the pre-only BMs in every NEM region. Eligibility would never exceed 10% under the pre-only BMs and would range as low as about 3.5%. By contrast, under the pre/post BMs eligibility would range from about 7.5% to over 15%.
- There is not a material spread of the eligible proportion across the pre-only or the pre/post BMs ***within*** each of the NEM regions, though there is a material difference ***across*** the regions, including:
  - A significantly higher proportion of the NMIs in Victoria and NSW than in other regions would be eligible under these thresholds
  - Tasmania's eligibility would be significantly lower than for other regions.

Table 25: Eligible percentage where the NMI is required to provide an RRMSE of 0.1 or less with 100% frequency – Large Commercial and Industrial Customers, 3:30-8:00pm event window, all quarters of 2019 by region

	Pre only adjustments				Pre and Post adjustments			
Region	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%
NSW	8.4	8.4	**9.4**	**9.4**	14.7	**15.0**	14.6	14.9
QLD	5.5	5.4	6.3	**6.4**	11.4	**11.6**	10.9	11.3
SA	7.4	7.3	**8.4**	8.3	11.3	11.5	**11.6**	11.4
TAS	3.5	3.5	**3.7**	**3.7**	7.5	**7.9**	7.3	7.7
VIC	9.7	9.8	9.9	**9.9**	14.9	**15.3**	14.5	14.8

Table 26 provides similar information for the 0.2/100% threshold. It shows that:

- It would significantly increase eligibility:
  - Under the pre-only BMs eligibility would increase by factors ranging from about 3 in the case of Victoria to about 6 for Tasmania
  - Under the pre/post BMs the increases would be slightly less but still substantial – ranging from a factor of about 3 to 4.
- But, as noted above, their use would essentially double the expected level of error in the administration of the WDRM.

Table 26: Eligible percentage where the NMI is required to provide an RRMSE of 0.2 or less with 100% frequency – Large Commercial and Industrial Customers, 3:30-8:00pm event window, all quarters of 2019 by region

Region	Pre only adjustments				Pre and Post adjustments			
	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%
NSW	**33.7**	31.4	32.9	31.4	43.9	**44.0**	43.2	43.2
QLD	33.5	31.8	**36.1**	35.0	44.5	46.1	45.0	**46.3**
SA	30.0	28.3	**31.1**	29.6	**39.5**	39.4	39.2	39.1
TAS	23.7	24.3	**25.2**	22.8	31.9	**36.4**	31.6	35.6
VIC	**32.6**	30.7	32.2	31.2	**42.0**	41.9	41.6	41.9

Table 27 shows the same information for the 0.1/90% threshold, and shows results similar to those across the regions in that it would:

- Result in proportions of eligible NMIs in each region being much closer to those of the 0.2/100% threshold than the 0.1/100% threshold
- Eligible proportions would range from about 18% in Tasmania to just under 30% for NSW in the pre-only BMs and from just over 30% for Tasmania to about 38.5% for NSW under the pre/post BMs

Table 27: Eligible percentage where the NMI is required to provide an RRMSE of 0.1 or less with 90% frequency – Large Commercial and Industrial Customers, 3:30-8:00pm event window, all quarters of 2019 by region

Region	Pre only adjustments				Pre and Post adjustments			
	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%
NSW	2767	28.0	29.1	**29.3**	37.1	**38.4**	36.7	37.8
QLD	23.8	23.5	26.7	**26.8**	36.3	**37.0**	35.6	36.5
SA	25.8	25.8	**27.4**	**27.4**	32.9	33.5	32.8	**33.6**
TAS	17.9	17.5	**18.9**	18.5	30.3	33.5	30.5	**34.1**
VIC	27.3	**27.4**	27.1	27.2	36.3	**36.8**	35.9	36.2

Further tables that provide information about the proportion of NMIs that would be eligible in each region in each quarter of 2019 based on each of the three threshold levels are included in Appendix D:.

4.2. Short early evening event window (5:30-7:00pm)

4.2.1. Assessment of accuracy, bias and variability of BMs for Large Commercial and Industrial Customers

Table 28 below provides the results of this analysis for all 8 candidate BMs across all regions and quarters for which we received consumption information.

As can be seen:

- BMs with multiplicative adjustment factors always produce lower median RRMSEs than those with additive adjustment factors, and have median ARE scores well within the 'good' range
- BMs with pre- and post-adjustment periods always provide lower median RRMSEs than those with only pre-adjustment periods
- A 20% adjustment factor provided lower median RRMSEs across the three-year period for the pre-only BMs, though on a quarter by quarter basis the 40% factor produced marginally better median RRMSEs in Quarters 1 and 4.
- A 40% adjustment factor provided lower median RRMSEs in the pre/post BMs in all of the timeframes assessed.

Table 28: Comparison of BMs for all years and regions, by quarter - Large Commercial and Industrial Customers, 5:30-7:00pm event window

Regions	Years	Qtr	Outcomes	Comments on median RRMSE outcomes	Best = Multiplicative	
					Pre	Pre&Post
All	All	All	**RRMSE** medians: Max 0.059, Min 0.077. **ARE** medians: -0.001 to +0.001. **RER** median values are not significant. RERT equivalent 90th percentile: 0.433	Pre only: Multiplicative RRMSE better by 0.004, adjustment percentage no real difference. Pre and post RRMSE better by ≈ 0.015 with multiplicative better	0.073 20% -0.001	0.059 40% +0.001
All	All	1	**RRMSE** medians Max 0.082, Min 0.061. **ARE** medians: -0.003 to 0.0. **RER** median values are not significant. RERT equivalent 90th percentile 0.474	Pre only: Multiplicative RRMSE better by 0.007, Adjustment percentage of 40% marginally better. Pre and post RRMSE better by 0.013 to 0.019 with multiplicative and 40% caps better	0.074 40% -0.002	0.061 40% -0.001
All	All	2	**RRMSE** medians Max 0.073, Min 0.056. **ARE** medians: +0.001 to +0.002. **RER** median values are not significant. RERT 90th percentile 0.403	Pre only Multiplicative RRMSE .002 better, with no difference for adjustments. All pre and post RRMSE around the same with the higher cap marginally better.	0.071 20% -0.003	0.056 40% -0.001
All	All	3	**RRMSE** medians Max .069, Min .054. **ARE** medians: +0.001 to +0.002. **RER** median values are not significant. RERT 90th percentile 0.385	All pre only BMs around the same RRMSE ≈ 0.068. All pre and post RRMSE around the same ≈ 0.055.	0.068 20% +0.001	0.054 40% +0.002
All	All	4	**RRMSE** medians Max .084, Min .064. **ARE** medians: +0.002 to +0.003. **RER** median values are not significant. RERT 90th percentile 0.479.	RRMSE for multiplicative better for pre only adjustment by ≈ .005 with the 40% adjustment better again. Pre and post multiplicative with 40% cap is best.	0.078 40% +0.002	0.064 40% +0.003

Table 29 provides similar information for the 8 candidate BMs across all regions for all of 2019 and each quarter within it. It shows that:

- Multiplicative adjustment factors produce lower median RRMSEs than those with additive adjustment factors in all cases for the pre-only BMs. This was also true for the pre/post BMs except for Quarter 3, in which an additive adjustment with a 40% cap provided a marginally better median RRMSE.
- The median ARE scores of all BMs were within the 'good' range.
- BMs with pre- and post-adjustment periods provide median RRMSEs that are equal to or lower than those with only pre-adjustment periods in all time periods except Quarters 3 and 4 where they are just marginally higher.
- A 20% adjustment factor provided lower median RRMSEs across the full time period considered and for all of the quarters except Quarter 1
- The 40% adjustment factor provided lower median RRMSEs in all quarters for the pre/post adjustment BMs.

Table 29: Comparison fof BMs or all regions for 2019 by quarter – Large Commercial and Industrial Customers, 5:30-7:00pm event window

Regions	Years	Qtr	Outcomes	Comments on median RRMSE outcomes	Best = Multiplicative	
					Pre	Pre&Post
All	2019	All	**RRMSE** medians Max 0.079, Min 0.060. **ARE** medians: -0.001 to +0.001. **RER** median values are not significant. RERT 90th percentile 0.425.	Pre only: Multiplicative is ≈ 0.004 better with no material change for adjustment %. Pre and Post is better by ≈ 0.016 with additive and multiplicative equal* for the higher adjustment amounts.	0.075 20% -0.001	0.060* 40% +0.001
All	2019	1	**RRMSE** medians Max 0.083, Min 0.061. **ARE** medians: -0.003 to -0.001. **RER** median values are not significant. RERT 90th percentile 0.469.	Pre only multiplicative is .007 better with a marginal improvement for the increased adjustment amount. Pre and Post is better by .006 to .016 with improvement for the higher adjustment amounts.	0.075 40% -0.002	0.061 40% -0.001
All	2019	2	**RRMSE** medians Max .070, Min 0.055. **ARE** medians: +0.001 to +0.002. **RER** median values are not significant. RERT 90th percentile 0.402.	Pre only multiplicative is ≈.003 better with no improvement for the increased adjustment amount. Pre and Post is better by .017 with marginal improvement for the higher adjustment amounts.	0.073 20% -0.003	0.057 40% -0.001
All	2019	3	**RRMSE** medians Max 0.082, Min 0.071. **ARE** medians: 0.002. **RER** median values are not significant. RERT 90th percentile 0.376.	Pre only multiplicative with 20% adjustment is marginally better with the others equal. Pre and Post is better by ≈ 0.014 with additive* 40% marginally better than the others.	0.069 20% 0.001	0.054* 40% 0.002
All	2019	4	**RRMSE** medians Max 0.086, Min 0.065. **ARE** medians: +0.001 to +0.002. **RER** median values are not significant. RERT 90th percentile 0.454.	Pre only: multiplicative is ≈ 0.004 better with a no improvement for the increased adjustment amount. Pre and Post is better by 0.016 with some improvement for the higher adjustment amounts.	0.081 20% +0.001	0.065 40% +0.002

4.2.2. Assessment of eligibility of Large Industrial and Commercial Customers at various accuracy/frequency thresholds

Table 30 provides this information for all of 2019 across all regions. As can be seen:

- The 0.1/100% threshold would result in between about 4.5% and 5.0% of the NMIs in Segment F being eligible to participate in the WDRM under the pre-only adjustment BMs and between about 9.5% and 10.25% under the pre/post BMs

- The 0.2/100% threshold would increase eligibility by a factor of over 5 under the pre-only BMs and a factor of about 4 for the pre/post BMs, but would double the total potential level of error that could be encountered
- The 0.1/90% threshold would result in eligibility levels quite close to the 0.2/100% threshold level, but the total potential error associated with this threshold would not be known with certainty.

Table 30: Eligible percentage for different required frequencies of producing a threshold RRMSE – Large Commercial and Industrial Customers, 5:30-7:00pm event window, all quarters of 2019, NEM-wide

	Pre only adjustments				Pre and Post adjustments			
RRMSE (BM%)	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%
0.1 100%	445	4.5	**5.0**	**5.0**	9.6	**10.3**	9.6	10.2
0.2 100%	26.3	23.1	**28.9**	26.6	39.0	40.8	39.2	**41.4**
0.1 90%	20.6	20.1	**23.1**	22.9	36.1	38.4	36.3	**38.5**

Table 31 provides similar information for the 0.1/100% threshold for all quarters of 2019 for each of the regions. It shows that:

- The pre/post adjustment BMs would allow a materially greater proportion of NMIs to be eligible to participate in the WDRM than the pre-only BMs in every NEM region. Under the pre-only BMs eligibility would never exceed 10% in Queensland, South Australia or Tasmania (where eligibility would never even reach 5%), and only marginally exceed 10% in Victoria and in some cases New South Wales
- Under the pre/post BMs eligibility would exceed 10% in all NEM regions, reaching about 15% in Queensland and South Australia, and up to about 18% in New south Wales and Victoria.
- There is not a material spread of the eligible proportion across the pre-only or the pre/post BMs ***within*** most of the NEM regions (the exception being Tasmania where there is a spread of 4 percentage points between the additive 20% and multiplicative 40% pre/post BMs. However, there is a material difference ***across*** the regions, including:
  - A somewhat higher proportion of the NMIs in Victoria and NSW than in other regions would be eligible under these thresholds
  - Tasmania's eligibility would be noticeably lower than that for the other regions.

Table 31: Eligible percentage where the NMI is required to provide an RRMSE of 0.1 or less with 100% frequency – Large Commercial and Industrial Customers, 5:30-7:00pm event window, all quarters of 2019 by region

	Pre only adjustments				Pre and Post adjustments			
Region	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%
NSW	9.9	9.9	**10.7**	10.6	17.5	**18.2**	17.6	18.2
QLD	6.2	6.1	6.7	**6.8**	14.1	**14.6**	13.9	14.5
SA	8.4	8.3	**9.9**	9.8	14.8	**15.4**	14.8	15.2
TAS	3.8	3.9	4.6	**4.6**	10.4	12.9	11.9	**14.4**
VIC	10.6	**10.7**	10.5	10.6	18.0	18.9	18.3	**18.9**

Table 32 provides similar information for the 0.2/100% threshold. It shows that:

- It would significantly increase eligibility:
  - Under the pre-only BMs eligibility would increase by factors ranging from about 3 in the case of New South Wales, South Australia and Victoria to about 5 fir Queensland and 6 for Tasmania
  - Under the pre/post BMs the increases would be slightly less but still substantial – with the factor ranging between about 2.5 in Victoria to about 3 in each of the other regions.
- But, as noted above, their use would essentially double the total potential level of error in the administration of the WDRM.

Table 32: Eligible percentage where the NMI is required to provide an RRMSE of 0.2 or less with 100% frequency – Large Commercial and Industrial Customers, 5:30-7:00pm event window, all quarters of 2019 by region

Region	Pre only adjustments				Pre and Post adjustments			
	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%
NSW	**35.5**	33.6	34.2	33.0	46.9	**48.3**	46.8	48.1
QLD	31.7	30.1	**34.3**	33.5	45.0	**47.5**	45.2	47.5
SA	31.5	28.5	**32.8**	30.8	43.9	**45.0**	43.6	43.9
TAS	**26.1**	24.8	26.0	24.8	36.3	39.5	36.3	**39.9**
VIC	**32.9**	31.7	32.6	31.5	45.00	46.8	45.1	**46.6**

Table 33 shows the same information for the 0.1/90% threshold, and shows results similar to those across the regions in that it would:

- Result in proportions of eligible NMIs in each region being much closer to those of the 0.2/100% threshold than the 0.1/100% threshold
- Eligible proportions would range from about 20% in Tasmania to just over 30% for NSW in the pre-only BMs and from 39% in Queensland, South Australia and Tasmania to just over 44% in New South Wales and Victoria under the pre/post BMs.

Table 33: Eligible percentage where the NMI is required to provide an RRMSE of 0.1 or less with 90% frequency – Large Commercial and Industrial Customers, 5:30-7:00pm event window, all quarters of 2019 by region

Region	Pre only adjustments				Pre and Post adjustments			
	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%
NSW	28.9	29.3	29.8	**30.2**	42.9	**44.5**	42.8	**44.5**
QLD	25.0	24.8	28.5	**29.1**	39.9	41.9	39.7	**42.8**
SA	26.3	26.0	27.7	**27.8**	39.2	**40.9**	38.3	39.
TAS	20.9	**22.0**	20.4	21.0	38.7	41.8	40.0	**43.1**
VIC	28.9	**28.9**	28.2	28.1	43.0	44.0	43.1	**44.1**

Further tables that provide information about the proportion of NMIs that would be eligible in each region in each quarter of 2019 based on each of the three threshold levels are included in Appendix D:.

4.3. Morning event window (7:00-9:00am)

4.3.1. Assessment of accuracy, bias and variability of BMs for Large Commercial and Industrial Customers

Table 34 below provides the results of this analysis for all 8 candidate BMs across all regions and quarters for which we received consumption information.

As can be seen:

- Pre-only BMs with additive adjustment factors always produce lower median RRMSEs than those with multiplicative adjustment factors; for the pe/post BMs the additive and multiplicative approaches produced very similar results
- All of the BMs exhibited median ARE scores well within the 'good' range
- BMs with pre- and post-adjustment periods always provide lower median RRMSEs than those with only pre-adjustment periods
- A 40% adjustment factor produced lower median RRMSEs for the pre-only BMs across the full timeframe and for Quarters 1 and 2, while the 40% cap produced better median RRMSEs in Quarters 3 and 4
- A 40% adjustment factor provided lower median RRMSEs in the pre/post adjustment BMs in all quarters.

Table 34: Comparison of BMs for all years and regions, by quarter – Large Commercial and Industrial Customers, 7:00-9:00am event window

Regions	Years	Qtr	Outcomes	Comments on median RRMSE outcomes	Best = Additive	
					Pre	Pre&Post
All	All	All	**RRMSE** medians: Max 0.070, Min 0.079. **ARE** medians: -0.001 to 0.0. **RER** median values are not significant. RERT equivalent 90th percentile: 0.363	Pre only additive* RRMSE better by ≈0.006, adjustment percentage no real difference. Pre and post RRMSE better by .008 to .014 with multiplicative equally good. Higher adjustment marginally better	0.078 40% 0.0	0.070 40% -0.001
All	All	1	**RRMSE** medians Max 0.086, Min 0.069. **ARE** medians: -0.003 to 0.0. **RER** median values are not significant. RERT equivalent 90th percentile 0.359	Pre only additive RRMSE better by 0.003, Adjustment percentage of 40% marginally better. Pre and post RRMSE better by 0.012 to 0.017 with multiplicative and additive around the same. 40% caps better	0.082 40% 0.0	0.069 40% -0.002
All	All	2	**RRMSE** medians Max 0.082, Min 0.071. **ARE** medians: -0.003 to -0.004. **RER** median values are not significant. RERT 90th percentile 0.372	Pre only additive RRMSE 0.006 better, with no difference for adjustments. All pre and post RRMSE around the same with the higher cap marginally better.	0.076 40% -0.003	0.071 40% -0.003
All	All	3	**RRMSE** medians Max .076, Min .069. **ARE** medians: +0.001. **RER** median values are not significant. RERT 90th percentile 0.343	Pre only additive BMs abetter by 0.007. No difference for adjustments. All pre and post RRMSE around the same ≈0.066.	0.069 20% +0.001	0.065 40% +0.001
All	All	4	**RRMSE** medians Max .093, Min .073. **ARE** medians: 0.0 to +0.001. **RER** median values are not significant. RERT 90th percentile 0.377.	Pre only additive RRMSE better by ≈0.005 with no difference for adjustments. Pre and post better by 0.014 to 0.018 with additive and multiplicative the same. 40% cap is best.	0.087 20% 0.0	0.073 40% +0.001

Table 35 provides similar information for the 8 candidate BMs across all regions for all of 2019 and each quarter within it. It shows that:

- Pre-only BMs with additive adjustment factors always produce lower median RRMSEs than those with multiplicative adjustment factors; for the pe/post BMs the additive and multiplicative approaches produced very similar results, except in Quarters 2 and 3 in which the additive adjustment produced marginally lower median RRMSEs
- Median ARE scores were within the 'good' range for all of the candidate BMs.
- BMs with pre- and post-adjustment periods always provide lower median RRMSEs than those with only pre-adjustment periods
- A 20% adjustment factor produced lower RRMSEs across the year as a whole and in Quarters 2 and 4 for the pre-only adjustment factor BMs, while the 40% factor provide lower median RRMSEs in Quarters 1 and 3
- A 40% adjustment factor provided lower median RRMSEs across the year as a whole and in all quarters for the pre/post adjustment BMs.

Table 35: Comparison of BMs for all regions for 2019 by quarter – Large Commercial and Industrial Customers, 7:00-9:00am event window

Regions	Years	Qtr	Outcomes	Comments on median RRMSE outcomes	Best = Additive	
					Pre	Pre&Post
All	2019	All	**RRMSE** medians Max 0.085, Min 0.071. **ARE** medians: -0.002 to 0.0. **RER** median values are not significant. RERT 90th percentile 0.359.	Pre only: additive is ≈ 0.005 better with no material change for adjustment %. Pre and Post is better by 0.007 to ≈ 0.012 with additive and multiplicative equal* for the higher adjustment amounts.	0.080 20% -0.001	0.071 40% -0.001
All	2019	1	**RRMSE** medians Max 0.085, Min 0.069. **ARE** medians: -0.003 to -0.002. **RER** median values are not significant. RERT 90th percentile 0.351.	Pre only additive is .003 better with a marginal improvement for the increased adjustment amount. Pre and Post is better by .003 to .006 with additive and multiplicative equal. 40% adjustment best.	0.082 40% -0.001	0.069 40% -0.003
All	2019	2	**RRMSE** medians Max .083, Min 0.072. **ARE** medians: -0.003 to -0.004. **RER** median values are not significant. RERT 90th percentile 0.369.	Pre only additive is 0.005 better with no improvement for the increased adjustment amount. Pre and Post is better by .01 with additive 40% adjustment best.	0.078 20% -0.003	0.0 72 40% -0.004
All	2019	3	**RRMSE** medians Max 0.76, Min 0.066. **ARE** medians: 0.001. **RER** median values are not significant. RERT 90th percentile 0.341.	Pre only additive with 40% adjustment is 0.005 better. Pre and Post is marginally better for additive and by ≈ 0.008 for multiplicative. Additive 40% is marginally better than the others.	0.069 40% 0.001	0.066 40% 0.001
All	2019	4	**RRMSE** medians Max 0.095, Min 0.077. **ARE** medians: 0.0 to +0.002. **RER** median values are not significant. RERT 90th percentile 0.375.	Pre only: additive is ≈ 0.005 better with a no improvement for the increased adjustment amount. Pre and Post is better by 0.011 to 0.018. Additive and multiplicative 40% are equally best.	0.090 20% 0.0	0.077 40% +0.001

4.3.2. Assessment of eligibility of Large Commercial and Industrial Customers at various accuracy/frequency thresholds

Table 36 provides this information for all of 2019 across all regions. As can be seen:

- The 0.1/100% threshold would result in between 6.5% and 8.2% of the NMIs in Segment F being eligible to participate in the WDRM under the pre-only adjustment BMs and between about 10.2% and 10.7% under the pre/post BMs
- The 0.2/100% threshold would approximately quadruple eligibility in both types of BMs, but would double the potential level of total error that could be experienced

- The 0.1/90% threshold would result in eligibility levels very similar to those expected under the 0.2/100% threshold level.

Table 36: Eligible percentage for different required frequencies of producing a threshold RRMSE – Large Commercial and Industrial Customers, 7:30-9:00am event window, all quarters of 2019, NEM-wide

RRMSE (BM%)	Pre only adjustments				Pre and Post adjustments			
	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%
0.1 100%	**8.2**	8.2	6.6	6.8	10.3	**10.7**	10.2	10.6
0.2 100%	**32.3**	31.5	27.6	26.6	37.8	37.9	38.2	**38.3**
0.1 90%	**29.8**	29.7	25.0	24.5	35.6	36.7	35.2	**37.3**

Table 37 provides similar information for the 0.1/100% threshold for all quarters of 2019 for each of the regions. It shows that:

- The pre/post adjustment BMs would allow a greater proportion of NMIs to be eligible to participate in the WDRM than the pre-only BMs in every NEM region, though the difference in each region is never more than about 3 percentage points. Eligibility would never exceed 10% in any region under the pre-only BMs and would range as low as about 3.5% in Queensland and Tasmania.
- By contrast, under the pre/post BMs eligibility would range from about 5.5% in Tasmania to about 10.5% in New South Wales
- There is not a material spread of the eligible proportion across the pre-only or the pre/post BMs ***within*** each of the NEM regions, and though there is a difference ***across*** the regions:
  - With New South Wales, Victoria and Queensland showing higher eligibility under both types of BMs, but the difference between the regions with higher and lower eligibility proportions being smaller than that observed in Segment E and less than observed for Segment F in the other event windows
  - Tasmania's eligibility is lower than for other regions, but less so than in Segment E and in other event windows for Segment F.

Table 37: Eligible percentage where the NMI is required to provide an RRMSE of 0.1 or less with 100% frequency – Large Commercial and Industrial Customers, 7:30-9:00am event window, all quarters of 2019 by region

Region	Pre only adjustments				Pre and Post adjustments			
	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%
NSW	**8.2**	8.2	6.6	6.8	10.3	**10.7**	10.2	10.6
QLD	3.7	**3.9**	3.3	3.2	6.9	**7.1**	6.6	6.8
SA	**7.0**	**7.0**	5.4	5.3	8.2	**8.2**	7.4	7.5
TAS	3.6	3.6	**4.6**	3.6	5.5	**5.8**	5.3	5.5
VIC	**7.9**	7.8	5.6	5.7	9.9	**10.2**	9.6	9.9

Table 38 provides similar information for the 0.2/100% threshold. It shows that:

- It would significantly increase eligibility:
  - Under the pre-only BMs eligibility would increase by factors ranging from about 3 for some of the BMs as applied in South Australia to over 6 for some BMs in Queensland and approaching 7 for some BMs in case of Tasmania
  - Under the pre/post BMs the increases would be slightly less but still substantial – ranging from a factor of about 3.5 in for New South Wales and Victoria to over 5 in Queensland and Tasmania.
- But, as noted above, the use of the 0.2/100% threshold would essentially double the total potential level of error experienced in the administration of the WDRM.

Table 38: Eligible percentage where the NMI is required to provide an RRMSE of 0.2 or less with 100% frequency – Large Commercial and Industrial Customers, 5:30-7:00pm event window, all quarters of 2019 by region

Region	Pre only adjustments				Pre and Post adjustments			
	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%
NSW	**32.3**	31.5	27.3	26.6	35.9	37.0	36.0	**37.4**
QLD	**27.0**	26.6	24.6	23.3	37.8	37.9	38.2	**38.3**
SA	**32.2**	31.0	27.0	25.1	33.4	33.4	**33.8**	33.4
TAS	23.7	**27.7**	25.1	26.0	27.9	**30.8**	28.5	30.2
VIC	**31.9**	31.3	27.6	25.6	36.2	36.4	36.1	**36.7**

Table 39 shows the same information for the 0.1/90% threshold, and shows results similar to those across the regions in that it would:

- Result in proportions of eligible NMIs in each region being much closer to those of the 0.2/100% threshold than the 0.1/100% threshold
- Eligible proportions would range from about 23% in Tasmania to just under 30% for New South Wales, South Australia and Victoria under some of the pre-only BMs, and under the pre/post BMs from just over 25% for Tasmania to as high as 37% for NSW under one of these BMs.

Table 39: Eligible percentage where the NMI is required to provide an RRMSE of 0.1 or less with 90% frequency – Large Commercial and Industrial Customers, 5:30-7:00pm event window, all quarters of 2019 by region

Region	Pre only adjustments				Pre and Post adjustments			
	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%	Additive 20%	Additive 40%	Multiplicative 20%	Multiplicative 40%
NSW	**28.7**	28.7	24.3	24.1	34.9	36.7	35.2	37.3
QLD	**24.4**	24.3	18.9	19.0	32.9	33.9	33.1	**33.8**
SA	**28.1**	28.1	21.5	21.9	31.0	**31.4**	30.0	30.5
TAS	24.3	**25.4**	22.8	22.9	25.7	27.6	26.1	**28.2**
VIC	**29.8**	29.7	25.0	25.0	35.6	**36.1**	34.7	35.4

5. Very Large Industrial Customers

5.1. The nature of this segment

This segment G contains customers with loads over 100GWh per year. There are only 21 of these customers in the NEM. Because of this very small number of customers, the analysis of this segment was d conducted on a site basis.

Of the 21 sites in this segment, five customers show a nett export from their sites during some, if not all, quarters, which renders analysis of those site meaningless for the Phase 1 study. These sites[30] are numbers 33470, 26283, 40090, 47220, 44507 and have been excluded from the BM analysis presented below.

The analysis therefore covers the remaining 16 sites; it assessed the median RRMSE and ARE that each of the BMs would produce when applied to each of these sites. Analysis results revealed that:

- Some sites have very stable loads that provide good median RRMSE results under all of the BMs. For example, site 7184 shows a median RRMSE (2019, all quarters) of ≈0.09 with a 90th percentile range of ≈0.05 for all BM methods. The median ARE varies from +0.001 (pre only) to -0.001 (pre and post).
- Other sites are not amenable to the use of any of the BMs. For example, site 11346 shows a median RRMSE of between 0.103 and 0.174 with 90th percentiles as high as 0.7 in all quarters in 2019. Its median ARE ranges from -0.33 to -0.54 with a 90th percentile greater than +0.4 in most cases.

The variation across customers is consistent with the experience of retailers with very large customers. Contracts for customers of this size are tailored to meet the specific requirements of the customer, combining the loads and any generation at the site to optimise the outcome for the customer. Similarly, very large sites would normally be treated individually for the measurement and use of DR.

5.2. Assessment of accuracy, bias and variability of BMs for Very Large Industrial Customers

The comparison of the BM method for each site indicated that those sites that qualify with acceptable median RRMSE and median ARE scores are relatively indifferent to the BM method employed, except for differences between the pre-only and the pre/post BM approaches. The three tables below show the median RRMSE and median ARE scores for the best pre-only and pre/post BMs for each site for each event window. As can be seen, the differences in most cases are marginal. Outcomes outside the 'good' ranges are shown in red.

30 Site numbers 33470, 26283, 40090, 47220 and 44507. (Note these are randomly assigned numbers that can be traced back to NMIs only by AEMO.

Table 40: Best BM for Very Large Industrial Customers (by site) – Morning event window (7:00-9:00am), all quarters of 2019

Site Number	Pre-only median RRMSE/ARE	Pre/post median RRMSE/ARE	Comments
4336	0.048 0.0	0.040 -0.001	Pre: Additive 20% Pre & post: Additive 40%
7184	0.009 +0.001	0.008 -0.001	All methods have the same value
9732	0.023 +0.003	0.018 -0.001	Pre: Multiplicative 20% or 40% Pre & post: Additive 40%
10380	0.045 +0.002	0.051 -0.002	Pre: Multiplicative 20% or 40% Pre & post: Additive 40%0.176
11346	0.136 -0.051	0.103 -0.033	Pre: Multiplicative 20% or 40% Pre & post: Additive 40%
19393	0.178 +0.024	0.0152 +0.012	Some methods with higher RRMSE scores pass the ARE requirements
21607	0.045 +0.004	0.045 -0.002	Pre: Multiplicative 20% or 40% Pre & post: -0.002
25162	0.004 0.0	0.005 +0.001	Pre: All methods equal Pre & post: All methods equal
29145	0.031 +0.001	0.031 -0.002	Pre: Multiplicative 20% or 40% Pre & post: Additive 40%
29286	0.037 -0.006	0.032 -0.002	Pre: All methods equal Pre & post: All methods equal
32952	0.004 0.0	0.006 +0.001	Pre: All methods equal Pre & post: All methods equal
34726	0.0151 -0.055	0.080 -0.033	Pre: Multiplicative 20% or 40% Pre & post: Multiplicative 40%
42687	0.171 -0.060	0.109 -0.041	Pre: Multiplicative 40% Pre & post: Additive 40%
46663	0.042 +0.02	0.033 -0.004	Pre: Additive 40% Pre & post: Additive 20% or 40%
51358	0.023 +0.002	0.022 -0.002	Pre: Multiplicative 20% or 40% Pre & post: Multiplicative 40%
60197	0.0163 -0.074	-	* Wide range included negative values

Note that in many cases, pre-only is better than pre/post, which is different to the outcomes in the E and F segments. In addition, there are a number of relatively high median RRMSE scores and some unacceptable median ARE scores.

Table 41: Best BM for Very Large Industrial Customers (by site) – Short early evening event window (5:30-7:00pm), all quarters of 2019

Site Number	Pre-only median RRMSE/ARE	Pre/post median RRMSE/ARE	Comments
4336	0.043 -0.006	0.031 0.0	Pre: Additive 40% Pre & post: Multiplicative or Additive 40%
7184	0.010 +0.002	0.008 0.0	Pre: All methods have the same value Pre and post: All methods have the same value
9732	0.023 -0.006	0.013 -0.003	Pre: Multiplicative 20% Pre & post: Multiplicative 20% or 40%
10380	0.063 -0.016	0.035 -0.006	Pre: Additive 20% Pre & post: Multiplicative 40%
11346	0.078 -0.005	0.065 -0.006	Pre: Multiplicative 20% or 40% Pre & post: Multiplicative 20%
19393	0.170 -0.019	0.146 -0.002	Pre: Multiplicative 20% Pre & post: Additive 40%
21607	0.057 -0.003	0.028 -0.001	Pre: Additive 20% Pre & post: Multiplicative 40%
25162	0.006 +0.001	0.004 +0.001	Pre: All methods have the same value Pre and post: All methods have the same value
29145	0.039 -0.011	0.021 -0.001	Pre: Additive 20% Pre & post: 40%
29286	0.039 +0.006	0.027 0.0	Pre: Multiplicative 20% Pre & post: Additive and multiplicative 40%
32952	0.006 +0.001	0.004 +0.001	Pre: All methods have the same value Pre and post: All methods have the same value
34726	0.087 -0.026	0.070 -0.010	Pre: Multiplicative 20% - note extreme value range Pre and post: Multiplicative 40%
42687	0.087 -0.012	0.074 -0.005	Pre: Multiplicative 20% Pre & post: Multiplicative 20%
46663	0.044 -0.007	0.024 -0.003	Pre: Multiplicative 40% Pre & post: Multiplicative 20% or 40%
51358	0.027 -0.004	0.018 -0.001	Pre: Multiplicative 20% Pre & post: Multiplicative 20% or 40%
60197	0.099 -0.027	0.078 -0.010	Pre: multiplicative 20% - note extreme value range Pre & post: Additive 40%

For this event window, pre/post adjustments are always better than pre-only, but the difference is small. In some cases, the ARE values are unacceptable and there are instances of high ranges for the BM RRMSE. For this segment, this means that the site cannot provide DR reliably.

Table 42: Best BM for Very Large Industrial Customers (by site) – Afternoon/evening event window (3:30-8:00pm), all quarters of 2019

Site Number	Pre-only median RRMSE/ARE	Pre/post median RRMSE/ARE	Comments
4336	0.054 -0.003	0.042 0.0	Pre: Additive 20% Pre & post: Additive 40%
7184	0.012 +0.001	0.011 +0.001	Pre: All methods have the same value Pre and post: All methods have the same value
9732	0.025 -0.005	0.021 -0.001	Pre: Multiplicative 20% or 40% Pre & post: Multiplicative 40%
10380	0.076 -0.010	0.065 -0.001	Pre: Additive 40% Pre & post: Multiplicative 40%

Site Number	Pre-only median RRMSE/ARE	Pre/post median RRMSE/ARE	Comments
11346	0.102 -0.006	0.102 +0.009	Pre: Additive 20% Pre & post: Additive 20%
19393	0.226 +0.005	0.202 +0.023	Pre: Additive 40% Pre & post: Additive 40%
21607	0.065 -0.010	0.052 +0.002	Pre: Additive 40% Pre & post: Multiplicative 40%
25162	0.007 +0.002	0.007 +0.002	Pre: All methods have the same value Pre and post: All methods have the same value
29145	0.048 -0.009	0.039 +0.002	Pre: Multiplicative 40% Pre & post: Multiplicative 40%
29286	0.047 +0.006	0.037 0.0	Pre: Multiplicative 20% or 40% Pre & post: Multiplicative 20% or 40%
32952	0.007 +0.003	0.007 +0.003	Pre: All methods have the same value Pre and post: All methods have the same value
34726	0.106 -0.035	0.091 -0.010	Pre: Multiplicative 20% Pre & post: Multiplicative 20%
42687	0.118 -0.003	0.112 -0.003	Pre: Additive 20% Pre & post: Additive 20%
46663	0.050 -0.004	0.040 0.0	Pre: Multiplicative 20% Pre & post: Multiplicative 40%
51358	0.030 -0.004	0.027 0.0	Pre: Multiplicative 40% Pre & post: Multiplicative 20% or 40%
60197	0.118 -0.037	0.104 -0.013	Pre: Multiplicative 20% - note extreme range Pre & post: Multiplicative 20%

Again, in this event window some sites have unacceptable ARE values and in one case the site median RRMSE is outside of the acceptable range. Pre/post adjustments are superior in all cases.

5.3. Assessment of eligibility of Very Large Industrial Customers at various accuracy/frequency thresholds

The number of the Very Large Industrial customer sites that are able to meet each of the three RRMSE/frequency criteria levels considered is shown in Table 43. The outcomes in the table assume that each site could use the BM to which its load was shown to be most suited and are consistent for all quarters of all years, indicating that it is a robust outcome.

Table 43: Number of sites that could participate for varying levels of RRMSE standards

Median RRMSE (Frequency%)	Afternoon/evening (3:30-8:00pm)	Early evening (5:30-7:00pm)	Morning (7:00-9:00am)
0.1 100%	0	0	0
0.2 100%	0	1	1
0.1 90%	4	4	4

This table shows, somewhat counterintuitively, that only one of the sites within Segment G will be able to provide baselines with good or acceptable accuracy for any of the event windows at all times. One site is able to meet the 'acceptable' level of accuracy for two of the event windows, the exception being the TI 32-40 window which extends beyond normal business hours, though four would exhibit baselines that would make them eligible to participate in all of the event windows at a good level of accuracy if they only have to meet that level 90% of the time. It should be noted that this represents approximately 20% of the sites in this segment, which is below (but not materially below) the eligible percentage achieved in the other segments for this criterion level.

6. Conclusions and discussion

This section summarises the findings presented in Sections 8 through 5 above. It should be noted that the following discussion is limited to results for the pre-only BMs as, in consultation with AEMO, it was decided that any pre/post BM approach would entail an obvious source of potential gaming. That is, the WDRU would be able to alter its consumption in the post-event adjustment period in ways that could affect the calculation of the baseline and therefore the amount of DR having been delivered. The decision was made based on the ***potential*** for such gaming to take place without consideration of the likelihood of it actually taking place[31].

6.1. Medium-Sized (160-750MWhpa) Customers

6.1.1. Preferred BMs overall and important variations

Table 44 below summarises the preferred BM for each event window for Medium-Sized Businee Customers, based on the ability of the BM to provide the lowest median RRMSE with an acceptable range of RRMSEs and a median ARE within the 'good' range. It shows that:

- A multiplicative adjustment with a 20% cap is the preferred choice for both the afternoon/evening (3:30-8:00pm) and the short early evening (5:30-7:00pm) event windows when the full year is considered; however, the 40% adjustment cap produces lower median RRMSEs in Q1 in all years
- By contrast, an additive adjustment with a 20% adjustment cap would produce the lowest median RRMSEs in the morning (7:00-9:00am) event window.

Table 44: Preferred and 2nd best BMs and associated median RRMSEs and median AREs – Medium-Sized Business Customers

Event window	Timeframe	Preferred (pre-only)	Median RRMSE	Median ARE	Next best or variation	Reduction in median RRMSE
3:30 to 8:00 PM	2017-2019	Multiplicative 20%	0.139	-0.002 to +0.001	40% cap better in Q1	Marginal
	2019	Multiplicative 20%	0.141	-0.003 to +0.001	40% cap better in Q1	-0.016
5:30 to 7:00 PM	2017-2019	Multiplicative 20%	0.123	-0.002 to +0.002	40% cap better in Q1	Marginal
	2019	Multiplicative 20%	0.124	-0.002 to +0.002.	None	No clear 2nd best
7:00 to 9:00 AM	2017-2019	Additive 20% or 40%	0.123	-0.001 to +0.002.	20% cap better in Q3 & Q4	Marginal in both
	2019	Additive 40%	0.123	-0.001 to +0.001	20% cap better in Q4	Marginal

31 Gaming in the post-event adjustment period can be mitigated to some extent by when that period starts and how long it extends. Conceptually, a longer duration, in particular, should impose a higher cost to be incurred by the WRDU in order to affect the amount of DR delivered against the baseline. However, no analysis of the likely impact of different durations on DR outcomes or the accuracy or bias of the BM itself was undertaken.

It is worth noting that the median RRMSE of the preferred BM in each event window falls outside the 'good' range of accuracy – despite that BM having been selected based on it exhibiting the lowest median RRMSE across the candidates. This means that more than half of all the baselines produced within the Medium-Sized Business Customer segment for each of the event windows failed to provide a level of accuracy of ±10% or less.

6.1.2. Expected eligibility at different accuracy/frequency thresholds for the preferred BM

Table 45 shows, for each of the event windows, the expected eligibility of Medium-Sized Business Customers under the three combinations of the accuracy and the proportion of the baselines in which that level of accuracy would need to be achieved in order for a candidate WDRU to be eligible to participate in the WDRM.

In each case, a preferred BM for the segment across all three event windows is used, with priority given to the preferred BM for the afternoon/evening event window (3:30-8:00pm) based on the fact that it contains the greatest proportion of high price events.

Table 45: Expected eligible percentage of Medium-Sized Business Customers, by region and threshold

Event Window	Preferred BM (based on 2019)	Threshold	NSW	QLD	SA	TAS	VIC
3:30 to 8:00 PM	Multiplicative 20%	0.1/100%	3.74	2.06	1.46	2.68	3.43
		0.2/100%	17.71	20.02	13.82	12.75	18.68
		0.1/90%	12.47	11.96	7.31	6.71	11.09
5:30 to 7:00 PM	Multiplicative 20%	0.1/100%	4.99	2.28	2.05	2.68	3.35
		0.2/100%	17.91	16.83	11.18	11.64	16.52
		0.1/90%	13.04	12.62	7.89	6.71	11.41
7:00 to 9:00 AM	*Multiplicative 20%* (see Note 1)	0.1/100%	3.15	1.6	0.88	1.79	2.08
			-0.54	0	0	0	-0.72
		0.2/100%	14.4	14.14	8.24	7.14	12.55
			-2.85	-1.65	-1.99	-4.26	-2.48
		0.1/90%	9.88	6.96	5.56	3.57	7.43
			-2.43	-1.77	-1.75	-1.69	-2.4

Note 1: Additive 40% was the preferred BM for this event window. For each of the 3 thresholds in this event window the first row shows the eligible percentage for the Multiplicative 20% BM (as the preferred BM for the segment) and the 2nd row shows the change in eligibility that would occur if the Additive 40% were used instead.

As can be seen:

- The 0.1/100% threshold would be expected to result in eligibility never reaching 5% for any of the event windows
- The 0.2/100% threshold would dramatically increase eligibility in all three event windows, with eligibility generally ranging between 10% and 20%, except in the case of the morning event window in which eligibility would still remain below 10% in South Australia and Tasmania
- The 0.1/190% threshold would produce eligibility levels closer to, but somewhat below those in the 0.2/100% threshold.

6.2. Large Commercial and Industrial Customers (750MWhpa to 100GWhpa)

6.2.1. Preferred BMs overall and important variations

Table 46 below summarises the preferred BM for each event window for Large Commercial and Industrial Customers, based on the ability of the BM to provide the lowest median RRMSE with an acceptable range of RRMSEs and a median ARE within the 'good' range. It shows that:

- A multiplicative adjustment with a 20% cap is the preferred choice for both the afternoon/evening (3:30-8:00pm) and short early evening (5:30-7:00pm) event windows when the consumption information available across all quarters is considered. However, the multiplicative adjustment with a 40% cap was the preferred BM when consumption from only 2019 is considered.
- An additive adjustment with a 20% cap is the preferred BM for the morning (7:00-9:00am) event window.

Table 46: Preferred and 2nd best BM and associated median RRMSE and median AREs – Large Commercial and Industrial Customers

Event window (TIs)	Timeframe	Preferred (pre-only)	Median RRMSE	Median ARE	Next best or variation	Reduction in median RRMSE
3:30 to 8:00 PM	2017-2019	Multiplicative 20%	0.085	-0.001 to +0.001	40% cap better in Q1	Marginal
	2019	Multiplicative 40%	0.087	-0.001 to +0.001	20% cap better in Q2 & Q3	Marginal in both
5:30 to 7:00 PM	2017-2019	Multiplicative 20%	0.073	-0.001 to +0.001	40% cap better in Q1	Marginal
	2019	Multiplicative 20%	0.075	-0.001 to +0.001	40% cap better in Q1	Marginal
7:00 to 9:00 AM	2017-2019	Additive 40%	0.078	-0.001 to +0.00	20% cap better in Q3 & Q4	Marginal in both
	2019	Additive 40%	0.080	-0.002 to +0.00	20% cap better in Q2 & Q3	Marginal in both

It is worth noting that the median RRMSE of the preferred BM in each event window falls within the 'good' range of accuracy. This means that more than half of all the baselines produced within the Large Commercial and Industrial Customer segment for each of the event windows provided a level of accuracy of ±10% or less.

6.2.2. Expected eligibility at different RRMSE and frequency levels

Table 47 shows the expected eligibility within the Large Commercial and Industrial Customer segment under the three combinations of the accuracy and the proportion of the baselines in which that level of accuracy needed to be achieved in order for a candidate WDRU to be eligible to participate in the WDRM with regard to each event window.

In each case, a preferred BM for the segment across all three event windows is used. Priority was given to the preferred BM for the afternoon/evening (3:30-8:00pm) event window, based on the fact that it contains the greatest proportion of high price events.

Table 47: Expected eligible percentage of Large Commercial and Industrial Customers, by region and threshold

Event Window	Preferred BM (based on 2019)	Threshold	NSW	QLD	SA	TAS	VIC
3:30 to 8:00 PM	Multiplicative 40%	0.1/100%	9.4	6.42	8.31	3.65	9.94
		0.2/100%	31.37	35.03	29.59	22.8	31.24
		0.1/90%	29.32	26.83	27.37	18.5	27.24
5:30 to 7:00 PM	*Multiplicative 40%*	0.1/100%	10.58	6.79	9.83	4.61	10.58
	(See note 1)		-0.12	0.07	-0.08	0.01	0.08
		0.2/100%	32.99	33.53	30.77	24.76	31.5
			-1.18	-0.76	-2.06	-1.25	-1.13
		0.1/90%	30.19	29.07	27.76	21	28.14
			0.37	0.54	0.08	0.58	-0.04
7:00 to 9:00 AM	*Multiplicative 40%*	0.1/100%	6.81	3.16	5.31	3.63	5.7
	(See note 2)		-1.36	-0.77	-1.66	0	-2.11
		0.2/100%	26.6	23.29	25.12	26	25.6
			-4.88	-3.3	-5.86	-1.72	-5.76
		0.1/90%	24.09	19.01	21.87	22.94	24.96
			-4.63	-5.29	-6.18	-2.49	-4.7

Note 1: Multiplicative 20% was the preferred BM for this event window. For each of the 3 thresholds in this event window the first row shows the eligible percentage for the Multiplicative 40% BM (as the preferred BM for the segment) and the 2nd row shows the change in eligibility that would occur if the Multiplicative 20% were used instead.

Note 2: Additive 40% was the preferred BM for this event window. For each of the 3 thresholds in this event window the first row shows the eligible percentage for the Multiplicative 40% BM (as the preferred BM for the segment) and the 2nd row shows the change in eligibility that would occur if the Additive 40% were used instead.

It should be noted that the preferred BM for the Large Commercial and Industrial Customer segment differs from that shown above for Medium-Sized Business Customers. AEMO may want to consider further the acceptability of the use of different BMs for different customer segments.

6.3. Very Large Industrial Customers (more than 100GWhpa)

This segment is comprised of 21 very large industrial customers with loads over 100GWh per year. The analysis of this segment is therefore different and was conducted on a site basis.

Of the 21 sites in this segment, five customers show a nett export from their sites during some, if not all, quarters, which renders analysis of those site meaningless for the Phase 1 study. The analysis therefore covered the remaining 16 sites. It showed that:

- Some sites have very stable loads that provided 'good' results under all of the BMs
- But for other sites, 'good' results did not occur under any of the candidate BMs.

The variation across customers is consistent with the experience of retailers that serve these sorts of customers. Contracts for customers of this size are tailored to meet the specific requirements of the customer, combining the loads, and any generation at the site, to optimise the outcome for the customer. Similarly, sites of this size would normally be treated individually for the measurement and use of DR.

Table 48 shows the number of sites within this segment that would be able to meet the three RRMSE/frequency criteria levels assuming that each could use the BM to which its load was shown to be most suited. The outcomes in the table are consistent for all quarters of all years, indicating that it is a robust outcome.

Table 48: Number of sites that could participate for varying levels of RRMSE standards

Median RRMSE (Frequency%)	Afternoon/evening (3:30-8:00pm)	Early evening (5:30-7:00pm)	Morning (7:00-9:00am)
0.1 100%	0	0	0
0.2 100%	0	1	1
0.1 90%	4	4	4

This table shows, somewhat counterintuitively, that none of the Very Large Industrial Customer sites are expected to be able to provide baselines with good or acceptable accuracy for any of the event windows at all times. One site is able to meet the 'acceptable' level of accuracy for two of the event windows, the exception being the afternoon/evening (3:30-8:00pm) window which extends beyond normal business hours. However, four sites would be expected to exhibit baselines that would make them eligible to participate in all of the event windows at a good level of accuracy if they only have to meet that level 90% of the time. It should be noted that this represents approximately 20% of the sites in this segment, which is below (but not materially below) the eligible percentage achieved in the other segments for this criterion level.

6.4. Recommendation of 'best' BM overall

Table 49 on the following page provides information on the 'best' BM in each customer segment based on the results presented in the Sections 4 through 6, where 'best' is defined as the BM that produces the lowest median RRMSE and a median ARE within the acceptable range[32].

32 The test for the 'best' BM was undertaken on a quarter by quarter and an annual basis and across the full 2017-2019 timeframe for each customer segment in each NEM region.

Table 49: 'Best' BM by customer segment

Segment	Best BM (lowest median RRMSE)	Comments
Medium-sized business customers (160 to 750 MWhpa)	Multiplicative adjustment with 20% cap	• Produces a median RRMSE in the 'acceptable' range in both the afternoon/evening and short early evening event windows across the 2017-2019 period as a whole • ARE median values are in the 'good' range • A multiplicative adjustment with a 40% adjustment cap produces marginally better median RRMSEs in Q1 in all years • The additive adjustment with a 20% cap produces the lowest RRMSEs in the morning event window
Large commercial and industrial customers (750 MWhpa to 100 GWhpa)	Multiplicative adjustment with 40% cap	• The multiplicative adjustment with a 20% cap provides 'good' median RRMSE in both the afternoon/evening and short early evening event windows across the 2017-2019 period as a whole • ARE median values are in the 'good' range • However, the multiplicative adjustment with a 40% adjustment cap produces marginally better median RRMSEs in Q1 in all years • The additive adjustment with a 40% cap produces the lowest RRMSEs in the morning event window
Very large industrial customers (Over 100 GWhpa)	None – recommendation is to consider allowing site-specific BMs	• Analysis was undertaken at the NMI level rather than across NMIs due to the small number of customers in this segment • Five of the 21 NMIs in this segment showed nett export during some, and in some cases, all quarters, which eliminated them from further consideration • In each of the other 16 NMIs, at least one of the BMs provided median RRMSE scores in the 'good' range in each of the event windows (with he exception of one NMI in the afternoon/evening event window) • But none of the BMs proved to be 'best' across a significant number of the 16 sites • Several sites were able to provide only 'acceptable' median ARE scores in each of the event windows

As noted in the table above, the best' BM differs by customer segment and time of day:

- For both Medium-Sized Business Customers and Large Commercial and Industrial Customers, a multiplicative adjustment approach was shown to be better in the afternoon/evening and short early evening event windows, while an additive approach proved better in the morning event window.
- But a different adjustment cap was seen to perform better in those two customer segments. A 20% cap proved better for Medium-Sized Business Customers while the 40% cap was better for Large Commercial and Industrial Customers.
- No single BM was shown to consistently provide the best results in among Very Large Industrial Customers.

However, from an administrative perspective, as noted in section 2.6.4 above:

- Ideally, the threshold level of accuracy and frequency to be applied should not vary by region, segment or season. Rather the selection of these thresholds should reflect a balance between what is deemed to be an acceptable level of error (bounded by the accuracy requirements laid out in the Rule change) and a reasonable level of customer eligibility.
- In particular, the use of the same BM in all regions and for all customer segments is likely to be preferable for the perceived equity of the WDRM as well as for administrative simplicity. However, where different BMs can be expected to be able to produce a higher level of eligibility at the selected level of accuracy when applied to different regions or customer segments, a case could be made for the applicable BM to vary. From an administrative perspective such a variation would be easier to implement regionally than by customer segment. Variation by season or event window could be very difficult operationally and is not recommended.

Based on these considerations, if AEMO would prefer to use same BM in all regions and for all customer segments, our recommendation would be to:

- Use the multiplicative adjustment BM with a 40% cap on the adjustment for both the Medium-Sized Business Customer and the Large Commercial and Industrial Customer segments in all time periods in all seasons and NEM regions. Our reasons are as follow:
  - The multiplicative approach is preferred because the afternoon and evening event windows include more high price events than any other parts of the day, including the morning event window
  - The 40% adjustment cap proved better for the Large Commercial and Industrial Customer segment which can be expected to have greater DR potential than the Medium-Sized Business Customer segment.
- Consider allowing Very Large Industrial Customers that want to participate in the WDRM to use one of the other pre-period only adjustment BMs where that BM provides a lower median RRMSE, particularly where that alternative will provide a median RRMSE that qualifies the site for participation when the multiplicative BM with 40% adjustment cap would not. We note that this would:
  - Increase the amount of DR that could be offered into the WDRM
  - Provide the first step in considering an expansion of the BMs used in the WDRM.

This approach, we believe, would provide a BM for the WDRM that:

- Is simple for customers and DRSPs to understand
- Reduces administrative burden for customers, DRSPs and AEMO
- Maintains a proper focus on accuracy in line with the AEMC's Rule change, but balances that with both simplicity and the desire to provide a reasonable level of eligibility for customers
- Provides the opportunity for further evolution and sophistication in the BMs used in the WDRM.

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: Functional specification

Est. 2008

Oakley Greenwood

WDR Baseline Method Testing

Functional Specification

prepared for:
AEMO

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

DOCUMENT INFORMATION

Project	Functional Specification
Client	AEMO
Status	Issued
Report prepared by	Angus Rich
Date	28 October 2020

Revision	Date	Modified by	Description
0	14/9/20	AR/LH	Initial Issue
1	21/9/20	AR	p4. Inclusion of negative interval treatment p9. Discussion point on no load intervals and periods
2	28/10/20	AR/LH	Included agreed positions on key questions in the Functional Specification. Updated statistical formula for adjustment factors and RER calculation. Included weather station locations.

1. Introduction

This document provides the functional specification for the calculation and procedures to process the data, using R script, for the Baseline Methodology Testing (BLM).

The data handling and calculation will be in three stages:

- Stage 1: Data transformation from AEMO supplied datafiles to datafile format required for stage 2.
- Stage 2: Processing of NMI data to a results database
- Stage 3: Datamining and testing of the results for the development of suitable acceptance criteria.

It is a live document and will be updated from time to time.

There are several discussion points identified through the document that need clarification and agreement between OGW and AEMO.

2. Stage 1: Data Transformation

The data input file for the R script will be flat files that will be assembled from a number of source files. The input file will consist of record that will be based on time and date order.

The key file will be the NMI files provided by AEMO using a daily record format described as follows:

Data field	Description
JURISDICTIONCODE	NSW, SA, VIC, TAS, QLD
MDM	MDM, MDM2
MWH_BAND	E. 160 to 750 MWh F. 750 MWh to 100 GWh G. 100+ GWh to 1000GWh H 1000+ GWh
NMI	NMI id
INTERVAL_STARTDT	Date of the record
NMICLASSCODE	SMALL, LARGE
COUNT_READS	48
SUM_VALUE01.....SUM_VALUE48	kWh for interval
MAX_VALUE_FLAG_AEZ	A (Actual), E (estimated by MDP) or Z (Estimated by AEMO)
DELIVERY_DELAY	Count of days between the read date, and the read delivery to AEMO date

Table 1 AEMO Segment NMI count for analysis

Annual consumption range	Meter Type	ACT	NSW	QLD	SA	TAS	VIC	Total NMI counts
1000+ GWh	COMMS1	0	2	3	0	1	1	7
100+ GWh to 1000 GWh	COMMS2	2	55	14	6	2	32	111
0.75+ GWh to 100 GWh	COMMS3	193	4,217	3,323	997	294	3,806	12,830
160+ MWh to 750 MWh	COMMS4	1,094	15,586	9,376	3,074	1,494	11,406	42,030

The range for the data supplied is 1/3/16 to 31/8/20.

It is expected that the data files will be 78 files,~58 GB in total.

Given the privacy and legal requirements of the NMI identification, OGW will generate an alias for the NMIs and provide a mapping file back to AEMO. For all results generated during this project all NMIs will be referenced by their alias. **For privacy protection, once the NMI map to OGW identifier is complete, it and the original datasets will be given back to AEMO; OGW will not retain a copy of either the original data or the map.** We assume AEMO will retain the map and the original data set for its own purposes and for the duration it deems necessary.

Additional datafiles required for the input file are:

- Regional node 30min wholesale price – each node, all dates in the data extract timeframe
- Regional node 30min wholesale demand - each node, all dates in the data extract timeframe
- NMIs in the DR portal (and date they were first listed in the portal if possible)
- NMIs contracted for RERT and dates and times on which they have been activated and responded
- Weather data for NEM cities
  - Archerfield (Queensland), Bankstown (New South Wales), Hobart Airport (AP) (Tasmania), Kent Town (South Australia), Melbourne Airport (AP) (Victoria), Melbourne Olympic Park (OP) (Victoria) and Penrith (New South Wales). These weather stations have the largest influence on demand forecasts for their respective NEM regions.
- Holidays for each region.

Data testing will be conducted on the AEMO dataset to ensure:

- Data is complete, no null data points etc.
- The segment classification will also be tested as it is understood that AEMO have allocated the segment by only testing October data only. OGW will apply reclassification as required.
- Number of intervals that have substitute data for a NMI are not excessive.
- Negative data fields for SUM_VALUE interval will be assumed to be a net export. NMIs with negative SUM_VALUE will still be included in the baseline calculations and analysis. The record for any day that has a negative or export interval will be tagged in the results file.

Status as at this revision

The data transformation has been completed.

All additional data files requested have been received except for the following:

- NMIs in the DR portal (and date they were first listed in the portal if possible)
- NMIs contracted for RERT and dates and times on which they have been activated and responded.

These would still be good to get but are not on the critical path.

3. Stage 2: NMI Data Analysis

The process flow is that the data input files will be processed through the script producing result files. The results files will contain fields that will allow datamining, processing and determining of key metrics and hurdle rates across the entire data set and by subgroups such as jurisdiction and consumption band.

Statistical calculations for precision, bias and variability are calculated will be calculated. Relative Root Mean Square Error (RRMSE) is used for precision; Average Relative Error (ARE) for bias; and Relative Error Ratio (RER) for variability.

The primary variable proposed is the RRMSE with secondary variables ARE and RER.

In addition to the primary variables other explanatory variables will also be included in the record. These might contain information such as average and peak regional price across event period, average or max daily temperature,

A record[1] will be included for each "event". Where multiple events occur in a day then multiple records will be record for each event resulting in multiple records in a day.

The planned approach for testing and certification of the Baseline Testing Methodology (BLM) is to conduct an "everyday" examination for all NMIs in the data set and will be based on a counterfactual measurement of statistics. This approach is so that a large amount of statistical data and the suitability for a NMI to apply the 10 of 10 baseline method under a majority of circumstances that a price event might be activated. Only using actual historical price events, say >$300/MWh, can yield only a small number of events for some regions with many events only half hour in duration.

The approach is:

- To nominate a standard event period/s that are likely to be seasonal for each regional node based on likely price events that may trigger a WDR event occurring.
- The different Baseline method tests discussed as follows will be applied to each NMI during these nominated everyday event periods to produce a set of statistical measures.

3.1. 10 of 10 Baseline Approach

The fundamental approach to baseline methodology is that a profile for a load during a particular time frame is calculated by averaging the consumption with the relevant trading intervals over a specified number of days to derive the unadjusted baseline. The unadjusted baseline is then adjusted by comparing the load in the pre- or pre- and post- period intervals of the event timeframe and comparing these to the unadjusted baseline to calculate an adjustment factor.

The 10 of 10 baseline approach is the current methodology used for the RERT and will be the only methodology used for this analysis. However, some variations that do not fundamentally change the methodology from the current approach will be included in the testing. Any variations will be based on publicly available information or information provided by AEMO. Any of these variations will be agreed to with AEMO.

1 A "record" in a database is a row of data. A "field" is an element in that "record"

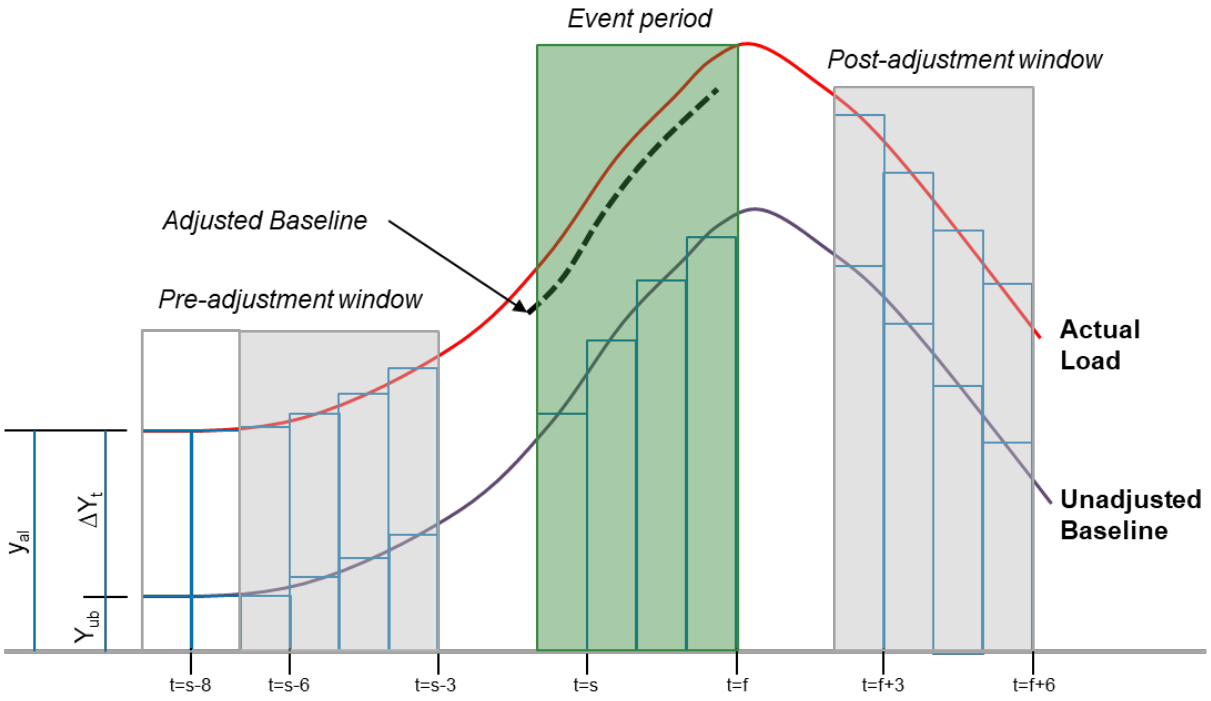
The baseline statistical measures will be calculated for each event for:

- 10 of 10 - no adjustment
- 10 of 10 - additive adjustment no cap
- 10 of 10 - additive adjustment with % cap
- 10 of 10 - multiplicative adjustment
- 10 of 10 - multiplicative adjustment with cap

3.1.1. Calculations

The calculations in this section are based on the AEMO RERT Panel Agreement for 2020-21. Figure 1 shows the basis of the pre-, and pre- and post- adjustment methodologies.

Figure 1: Baseline adjustment windows



s represents the start of the event period and f represents the end of the event period.

Unadjusted baseline calculation

$$b_t = \frac{1}{S}\sum_{i=1,2,..,S}^{n} c_{ti}$$

b = unadjusted baseline MWh for a given time interval (t)

i = one of S selected days

t = trading interval.

c = MWh electricity demand for a given trading interval (t) occurring on one of the selected days i.

The AEMO 2020-21 RERT Panel Agreement defines:

S = the set of selected days in the 100[2] calendar days immediately preceding the [day/weekday] on which reserve was activated and for which the calculation is being made (the 45 day period). The days in the 45 day period selected for the set will be based on [day/weekdays] on which reserve was not activated (Non-Activated Days) and [day/weekdays] on which reserve was activated (Activated Days) determined as follows:

Step 1 - This set of selected days will normally comprise the 10 Non-Activated Days immediately preceding the [day/weekday] on which reserve was activated and for which the calculation is being made.

Step 2 - If, in the 45 day period, there are less than 10 Non-Activated Days but 5 or more Non-Activated Days, then S comprises those Non-Activated Days.

Step 3 - If, in the 45 day period, there are less than 5 Non-Activated Days, then S comprises the Non-Activated Days plus one or more of the Activated Days in the 45 day period will added to the number of Non-Activated Days so that the total number of days in the set equals 5. The Activated Days added to the Non-Activated Days will be determined based on the level of demand during the trading intervals during the period of activation on the Activated Days (with the Activated Day with the highest demand during any trading interval during the period of activation on that Activated Day ranked highest and added to the Non-Activated Days, with the next highest ranked Activated Day added and so on, until the total number of days in the set equals 5). If 2 or more Activated Days are ranked the same based on the highest demand during any trading interval during the period of activation, the Activated Day closest in time to the [day/weekday] on which reserve was activated and for which the calculation is being made will be ranked higher.

For the OGW analysis of the everyday case counterfactual analysis in Phase 1 there is the assumption that there are no Activated Days. The only days excluded from the calculation are weekends and public holidays or as identified in the following sections.

In the case that a subset of accepted NMIs are processed through scenario testing in Phase 2 then Steps 1 to 3 will apply.

AEMO Discussion Point: How do we treat weekends?

It is unclear from the RERT Panel Agreement how weekends are treated as the reference is only made to [day/weekday] for the 45 day period (within 100 calendar days) to select the 10 days for calculating the unadjusted baseline. However, it does provide the opportunity for proponents to nominate reserve capacity for weekend periods.

OGW suggests adopting the CAISO 10 of 10 for weekends uses the average of 4 most recent weekend days out of the four most recent weekend days according to the work previously done by DNV-KEMA for AEMO in 2013.

OGW's suggestion was accepted; AEMO said they would be interested in our views on ways the weekend day test could be made more robust, including the possible use of a like-day approach.

2 It is unclear the reasoning behind the 100-day nomination and the 45-day period in the RERT Agreement.

AEMO Discussion Point: How do we treat public holidays?

The file provided shows only 7 common public holidays and not the typical 10 public holidays (e.g. Show day, Queen's birthday, Labour day) that are in each individual jurisdiction.

The RERT Agreement for the treatment of baselining states:

"weekday" means a day that is not a Saturday or Sunday or observed as a public holiday in the region in which the reserve is located.

The NER defines:

A business day is not a Saturday or Sunday; or observed as a public holiday on the same day in each of the participating jurisdictions (except the Commonwealth).

OGW seeks clarity on how public holidays shall be treated in the Wholesale Demand Response baseline methodology?

It was decided that because holidays will be characterised by different consumption patterns, they will be removed from the analysis using information on jurisdictional holidays provided by AEMO.

Adjustment calculations

The pre-adjustment only calculation[3] considers the window between interval t=s-3 to t=s-8 where s is the start of the event period.

The pre- and post-adjustment calculation considers the pre-adjustment window t=s-3 to t=s-6 and post-adjustment window of t=f+3 to t=f+6

Pre-Adjustment Window only

Additive Calculation

$$a_{adj} = \frac{\sum_{t=s-8}^{t=s-3}(c_t - b_t)}{6}$$

Multiplicative Calculation

$$m_{adj} = \frac{\frac{\sum_{t=s-8}^{t=s-3} c_t}{n}}{\frac{\sum_{t=s-8}^{t=s-3} b_t}{N}}$$

Pre- and Post- Adjustment Window Additive Calculation

Additive Calculation

[3] Current methodology for the RERT process only considers a pre-adjustment period.

$$a_{adj} = \frac{\sum_{t=s-6}^{t=s-3}(c_t - b_t) + \sum_{t=f+6}^{t=f+3}(c_t - b_t)}{8}$$

Multiplicative Calculation

$$m_{adj} = \frac{\frac{\sum_{t=s-6}^{t=s-3} c_t + \sum_{t=f+6}^{t=f+3} c_t}{N}}{\frac{\sum_{t=s-6}^{t=s-3} b_t + \sum_{t=f+6}^{t=f+3} b_t}{N}}$$

Where:

a = additive adjustment factor (this may be positive or negative)

m = multiplicative adjustment factor (1<m>1)

s = the start of the trading interval (t) during which the WDR has been activated and for which the calculation is being made.

c = MWh electricity demand for a given time interval (t) during the period of WDR for which the calculation is being made.

$s-n$ = trading interval n 30-min intervals before activation start time

Caps on Adjustment Factors

Traditionally caps on adjustment factors may have been introduced to prevent gaming by participants by artificially changing their loads during the adjustment windows to influence the adjustment factors. However, recent work by CAISO indicate that the traditional cap of 20% may be limiting in the circumstance of loads that tend to be weather dependent and have nominated a cap of 40%.

Adjustment factors will be treated symmetrically.

Additive caps

For additive caps any limitations to adjustments (increase or decrease) or caps are expressed as a percentage of the average of unadjusted baseline intervals that are during the adjustment windows.

That is for additive cap pre-only adjustment windows:

$$a_{adj\_cap} = cap\ (\%) \times \frac{\sum_{t=s-8}^{t=s-3} b_t}{N}$$

That is for additive cap pre- and post- adjustment windows:

$$a_{adj\_cap} = cap\ (\%) \times \frac{\sum_{t=s-6}^{t=s-3} b_t + \sum_{t=f+6}^{t=f+3} b_t}{N}$$

Multiplicative caps

For the multiplicative cap the ratio will be limited to 1 +/- cap. m_{adj} is calculated as above.

$$If\ m_{adj} > \big(1 + cap(\%)\big)\ then\ m_{adj\_cap} = (1 + cap(\%))$$

$$If\ m_{adj} < \big(1 + cap(\%)\big)\ then\ m_{adj\_cap} = (1 - cap(\%))$$

$$else\ m_{adj\_cap} = m_{adj}$$

Caps that will be implemented are:

- 20% - existing implied limitation of the RERT
- 40% - based on current analysis by Nexant for CAISO
- No cap

AEMO Discussion Point: Treatment of adjustment factors for multiple events in one day.

The current RERT Panel Agreement States:

If the reserve is activated for 2 or more separate periods on the same day, the adjustment factor for each period of activation will be the adjustment factor calculated for the first period of activation on that day.

OGW is proposing to adopt the same approach for the WDR baseline calculations.

AEMO Discussion Point: Days where no load no present

In some of the data files received there are records for some NMIs where the entire record for a day (all 48 fields) have zero kWh for the SUM_VALUES. It is OGW's assumption that the facility was offline for this day record. It was agreed that OGW will not calculate any baseline statistics for that day or for any event periods for which the aggregate of the SUM_VALUES is zero (the assumption being that in these periods there was no load available with which to participate in WDR).

However, OGW will include those days/periods in the calculation of any unadjusted baseline for dates that are after the day. This is in accordance with the fact that there are no provisions in the RERT Panel Agreement to exclude days on the basis of load type except for weekends, public holidays or activation days in the calculation of the unadjusted baseline. Flags will be used to identify days for which the baseline included zero consumption days (or zero-consumption event periods. The flag will allow us to determine whether the frequency at which these types of days/periods occur and the degree to which they affect the predictive power of the baseline.

3.1.2. Statistical Calculations

The statistical calculations described below will be applied and evaluated at the NMI level. The analysis will in essence create a set of baselines for simulated event periods on a large number of days using everyday analysis. Evaluation metrics will be applied to the set of baselines for each NMI. Final results will be provided regarding the number and total load of the NMIs meeting the evaluation thresholds for each BLM. As a starting point, aggregate results will be provided by BLM for each consumption segment and jurisdiction and other metrics and segmentations that may become relevant through the analysis.

Accuracy - Relative Root Mean Squared Error (RRMSE)

RRMSE is the primary measure and metric for the accuracy of the baseline

The RRMSE is calculated as follows:

$$RRMSE = \frac{\sqrt{\frac{\sum_{n \in N} (L_n^{baseline} - L_n^{actual})^2}{N}}}{\frac{1}{N} \times \sum_{n \in N} L_n^{actual}}$$

Where:

- n is the set of *trading intervals* from which *metering data* is taken for the of the calculation.
- *N* is the number of elements in set n
- $L_n^{baseline}$ is the calculated baseline load associated with a *trading interval* in set *n*.
- L_n^{actual} is the actual metered load associated with a *trading interval* in set *n*.

The RRMSE combines the systematic errors measured by the ARE (bias) and the variability of errors captured by the RER (variability). For this reason, the overall accuracy measure is given primary emphasis in this analysis.

Previous studies by KEMA for AEMO indicate that acceptance criteria for a successful BLM considered an RRMSE < 10% and was the reasoning for selecting the CAISO 10 of 10 with an additive adjustment as the methodology for the RERT[4].

Consequences of inaccurate baselines

> *When a baseline is "wrong" (i.e. it does not reflect what the consumer's electricity use would have been in the absence of demand response), it means that the quantity of demand response that was accounted for will be wrong. If the baseline is too high, the amount of demand response will be overestimated. If the baseline is too low, the amount of demand response will be underestimated. As a result, either too much or too little value relating to demand response will be transferred from the buyer to the seller of demand response. This will result in the DRSP either being paid for more demand response than was provided or being underpaid for the quantity provided.*
>
> *In a single instance, if the baseline is wrong, the demand response will either be over or undervalued. However, if the baseline is correct on average when wholesale demand response is being dispatched over time, then the fair value for the demand response should be exchanged between the retailer and the demand response provider. If it is correct on average, the over- and under-valuation of the demand response should cancel out over time.*
>
> *So, while in the short-term, the value attributed to demand response through settlements may be incorrect, the distortionary impacts should be at least partially mitigated in the medium-term if the average error in the baseline is zero[5].*

Bias – Absolute Relative Error

Average relative error (ARE), is a measure of bias derived by adding the differences between the data points in the baseline and the actual load points for each day.

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

5 AEMC, June 2020, NATIONAL ELECTRICITY AMENDMENT (WHOLESALE DEMAND RESPONSE MECHANISM) RULE 2020

$$ARE = \frac{\bar{b_t} - \bar{c_t}}{\bar{c_t}}$$

Where:

$\bar{b_t}$= average half hourly baseline during event period

$\bar{C_t}$= average half hourly actual meter measurement during event period

ARE of zero shows no bias, positive is over estimation, negative is under estimation.

The closer ARE is to zero, the closer the baseline is to being unbiased.

An unbiased result returns a median value of zero for all results of a NMI.

Variability – Relative Error Ratio

Previous studies by DNV KEMA for ARENA describe the measure of variability as Relative Error Ratio (RER). The definition described as the "standard deviation of the baseline's prediction errors expressed as a fraction of average load".

The standard deviation is the standard deviation of a sample (indicated by the (n-1))[6].

$$RER = \frac{\sqrt{\frac{\sum_{t=s}^{t=f}(b_t - c_t)^2}{n-1}}}{\bar{c_t}}$$

Where:

$(b_t - c_t)$ = predictive error of the baseline for an interval

n = number of half hour intervals in the event period

$\bar{C_t}$= average half hourly actual meter measurement during event period

The smaller the median RER for the NMI across the dataset analysed, the less variable a baseline's error is for the NMI and therefore the better the baseline performs across a wide variety of circumstances.

A median RER < 0.1 for a NMI is acceptable[7].

3.1.3. Candidate WDR Event periods

The previous work conducted by DNV-KEMA for testing different baseline methodologies testing identified that the top 100 half-hour demands for all regions occurred between 1pm to 5pm and between 6pm to 7pm in the months of October to March.

Their analysis was standardised on the period of 1pm-7pm as the event period for the calculation of everyday event statistics.

6 RER derived from Table 20 in *KEMA, 2014, NYISO SCR Baseline Study*

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

OGW is proposing to apply the same approach but to identify high price events instead of high demand events with the expectation that the periods are likely to reveal a morning and afternoon event period. The same event period/s will be applied across each day but may be examined on a regional and seasonal basis.

It is expected that the main price events will be in the Summer and Winter periods.

Candidate event periods will be defined on a regional basis.

Details to be provided and agreed upon.

The BLM parameters to be tested must be capable of including multiple windows per day over varying trading interval durations.

The event periods selected in consultation with AEMO are shown in Table 2. Tests 1 -3 are based on potential high price periods over the 2016-2020 time frame. The AEMO all day event was originally proposed to cover the periods 15-44 (ending 7:30am-10pm). For the purpose of processing and current coding design, this has been reduced to 15-42 to ensure that any post period adjustment window did not extend into the next day.

Table 2: Baseline event period test windows.

Test ID	Periods
Test 1 - morning peak	meter intervals 15-18 (ending 7:30am - 9am)
Test 2 - evening peak	meter intervals 36-38 (ending 6pm - 7pm)
Test 3 - AEMO all day	meter intervals 15-42 (ending 7:30am - 9pm)
Test 4 - afternoon high price	meter intervals 32-40 (ending 4pm - 8pm)

3.1.4. Key statistics

- For both pre-adjustment and pre- and post- adjustment baselines the following RRMSE, ARE, RER statistics will be calculated for:
  - 10 of 10 - no adjustment
  - 10 of 10 - additive adjustment no cap
  - 10 of 10 - additive adjustment with % cap
  - 10 of 10 - multiplicative adjustment no cap
  - 10 of 10 - multiplicative adjustment with cap

3.1.5. Other Statistical Data

For each NMI, statistical data will also be captured across **all time intervals** and the **event period/s only intervals** for the following load-defined segments:

- Calendar year load
  - Percentiles: 25%, 75%, 50% (median), percentile range, min, max
  - Load mean
  - Load factor (mean/max)
- Seasonal load

- Percentiles: 25%, 75%, 50% (median), percentile range, min, max
- Load mean
- Load factor (mean/max)

These additional NMI-level statistical measures will be used to explore whether there are any correlations or trends with the metrics and acceptance criteria defined through this project.

3.1.6. Weather Data

For each record the weather data for the day will be reported and will include min, max and average daily temperature as possible explanatory variables.

The weather stations used for data analysis are

- Kent Town (South Australia), Adelaide West Terrace (South Australia)[8]
- Archerfield (Queensland),
- Bankstown (New South Wales), Penrith (New South Wales)
- Hobart Airport (AP) (Tasmania)
- Melbourne Airport (AP) (Victoria), Melbourne Olympic Park (OP) (Victoria)

These weather stations have the largest influence on demand forecasts for their respective NEM regions[9].

3.1.7. Price Data

For each event period the minimum, maximum and average half hour price will be reported in each record.

3.1.8. Regional load data

For each event period the minimum, maximum and average regional demand will be reported in each record.

3.1.9. NMI and NMI meta-data

MAX_VALUE_FLAG_AEZ, COUNT_READS and DELIVERY_DELAY will be used to include or exclude NMIs or data sets from the analysis.

8 Adelaide West Terrace was not included in previous reports. Analysis at this station has been included due to the decommissioning of Kent Town by the Bureau of Meteorology on 31 July 2020. AEMO have now introduced Adelaide West Terrace into demand forecasting models in place of Kent Town.

9 AEMO, August 2020, *Temperature Forecast Analysis for Summer 2019-20.*

AEMO Discussion Point: Treatment of substitute data with baseline calculations?

The current RERT Panel Agreement states the following:

Measurement: have actual interval metering data (not estimated metering data or substituted metering data) for all trading intervals for at least 100 calendar days prior to [day/weekday] on which reserve was activated and for the period of activation.

OGW propose to exclude construction of a baseline and statistical calculations for any day on which substitute data exists in the NMI consumption record. However, substitute data for the preceding days use in the construction of the unadjusted baseline calculation will be deemed acceptable for event days that have (A) actual consumption data.

The decision was made to use estimated and substituted data in the analysis. This will apply to days for which baselines are created and days used in the creation of baselines. We will use flags to allow us to determine whether this approach makes a material difference to the outcome.

3.1.10. Other AEMO Data

For each record, other relevant AEMO data will also be reported (e.g., if the day is a RERT activation day).

3.1.11. Results Segmentation

Some of the statistics will be run across the data based on different segments (e.g. seasonal vs yearly) particularly statistics and data that may be influenced by seasonal outcomes with different load types.

Examples of statistics that may be used for defining segments of possible interest data segmented for each NMI:

- NMI load min, max, average, median, load factor, 25% and 75% quartile for both all intervals and event period intervals only.
- Percent of RRMSE less than acceptance hurdle 1, acceptance hurdle 2, etc.
- Median ARE and RER.

The purpose of these analyses will be to provide additional information regarding the degree to which different loads conform to the use of different BLMs and the application of different thresholds of acceptable accuracy, bias and variability.

4. Results Database

The intent is the results database will provide all the processed data for event as defined by the input data specifications. The results file will be datamined and summarised to meaningful views using R script to process subsets on the project team's parameters.

The process will be to run the results file back through the R script to be able to provide datamining, summaries of analysis and visualisation of the results.

As a starting point, data may be reviewed on these (or any combination):

- Regional node
- Load segment
- Annual and seasonal/quarter segmentation
- Load segmentation

Combinations of result analysis criteria will evolve through the process.

It will also be used to determine the metrics and parameters for what counts for success for the WDR program and develop the acceptance criteria based on the data reveal.

An example may be that a successful participant will achieve a high percentage of events (>90%) with an RRMSE < 10% and all events with an RRMSE < 20%. Actual metrics will emerge with the analysis of the event results and mutually agreed with AEMO.

For each NMI and each baseline calculation the results will calculated for the proportion of baselines that have RRMSE in the following ranges:

- +/-10% RRMSE
- +/-20% RRMSE
- +/- 30% RRMSE
- Outside the range of +/- 30% RRMSE

The results of this analysis will used to aggregate these results and report on the number of NMIs for each baseline that fall into the following categories and the starting point segmentations as above:

- For each RRMSE range: Time Compliance 100%, 95%, 90%, 85% compliance in 5% increments

Along side of the results segmentation statistical measurements will processed across the entire load segments that examine RRMSE, ARE, RER for each of the baseline calculation methodologies on the following basis:

- Average and standard deviation
- Median, and 10th and 90th percentiles

It is likely that the subset of NMIs that meet the everyday hurdle metrics may be further assessed by running scenario analysis. For example, actual historic defined event periods using a price threshold.

5. R script Details

1. An input set comprises a unique set of input settings. Each input is a variable. Many variables will not change between input sets, but the ability to change them if required is built into the testing process. They include
    a. NMI data qualifiers
    b. CAISO baseline parameters: candidate days (e.g. 45), near-like days (10); weekends (excluded), public holidays (excluded), RERT event days (excluded)
    c. Event day proxies are individually specified in a file. This allows for multiple event windows per day, and any variation of event duration, but not beyond midnight.
    d. One Window Adjustment parameters
        i. Additive, Additive Cap, Multiplicative, Multiplicative Cap,
        ii. Gap between Event and Adjustment Window
        iii. Size of Adjustment Window
    e. Two Window Adjustment parameters
        i. Additive, Additive Cap, Multiplicative, Multiplicative Cap,
        ii. Gap between Event and Prior Adjustment Window
        iii. Size of Prior Adjustment Window
        iv. Gap between Event and Post Adjustment Window
        v. Size of Post Adjustment Window (restricted to 00:00)
    f. Statistics
        i. RRMSE, ARE and RER for each adjustment method
        ii. Seasonal Quartiles for baselines
    g. Dates
        i. Analysis start date
        ii. Analysis end date (typically 1 year from Start Date)
2. Result Sets. A Result Set is associated with one Input Set and one NMI only. It comprises the all the Input Set parameters as columns, the NMI and a daily result of the Statistics.
3. Result Sets will be combined with NMI meta data and other data identified in the process into a table readily consumed by tools required for further analysis.
4. Further parameters may be added to Input Sets to create new ones and the process repeated.

Appendix C: Price event analysis results

Table 50: High price events -- all trading intervals

Quarter	Day type	NSW	QLD	SA	TAS	VIC
Q1	all days	204	254	363	156	247
	weekdays	168	164	299	121	209
	weekends	36	90	64	35	38
Q2	all days	70	26	100	39	69
	weekdays	66	26	93	31	64
	weekends	4	0	7	8	5
Q3	all days	111	10	146	41	171
	weekdays	94	8	132	30	152
	weekends	17	2	14	11	19
Q4	all days	60	19	74	49	78
	weekdays	57	16	66	41	74
	weekends	3	3	8	8	4

Table: 51: High price intervals in the Morning event window (7:00-9:00am)

Quarter	Day type	NSW	QLD	SA	TAS	VIC
Q1	all days	204	254	363	156	247
	weekdays	168	164	299	121	209
	weekends	36	90	64	35	38
Q2	all days	70	26	100	39	69
	weekdays	66	26	93	31	64
	weekends	4	0	7	8	5
Q3	all days	111	10	146	41	171
	weekdays	94	8	132	30	152
	weekends	17	2	14	11	19
Q4	all days	60	19	74	49	78
	weekdays	57	16	66	41	74
	weekends	3	3	8	8	4

Table 52 High price intervals in the Afternoon/evening event window (3:30-8:00pm)

		NSW	QLD	SA	TAS	VIC
Q1	all days	100	130	159	48	108
	weekdays	80	83	131	38	93
	weekends	20	47	28	10	15
Q2	all days	46	18	53	4	46
	weekdays	43	18	47	4	41
	weekends	3	0	6	0	5
Q3	all days	76	5	77	6	107
	weekdays	63	4	71	3	93
	weekends	13	1	6	3	14
Q4	all days	38	15	43	9	40
	weekdays	35	12	38	8	37
	weekends	3	3	5	1	3

Table 53: Duration of high price events (in contiguous trading intervals) 2019, by NEM region

State	Week segment	duration	all intervals	Test 01	Test 04	State	all intervals	Test 01	Test 04
NSW	alldays	1	135	18	79	TAS	157	25	27
		2	68	8	40		52	8	18
		3	36	0	33		12	0	3
		4+	206	0	108		64	8	19
	weekdays	1	118	18	65		121	23	20
		2	58	8	34		42	8	14
		3	30	0	30		6	0	0
		4+	179	0	92		54	8	19
	weekends	1	17	0	14		36	2	7
		2	10	0	6		10	0	4
		3	6	0	3		6	0	3
		4+	27	0	16		10	0	0
QLD	alldays	1	117	12	55	VIC	159	20	72
		2	46	4	18		86	12	46
		3	30	0	24		78	6	66
		4+	116	0	71		242	4	117
	weekdays	1	93	12	43		141	19	58
		2	36	4	16		76	12	36
		3	27	0	24		60	6	57
		4+	58	0	34		222	4	113
	weekends	1	24	0	12		18	1	14
		2	10	0	2		10	0	10
		3	3	0	0		18	0	9
		4+	58	0	37		20	0	4
SA	alldays	1	166	20	65				
		2	114	8	66				
		3	90	15	60				
		4+	313	4	141				
	weekdays	1	142	18	53				
		2	90	8	52				
		3	81	15	54				
		4+	277	4	128				
	weekends	1	24	2	12				
		2	24	0	14				
		3	9	0	6				
		4+	36	0	13				

Figure 7: Proportion of all high price intervals that occur the Morning (7:00-9:00am) or Afternoon/evening (3:30-8:00pm) event windows, 2019

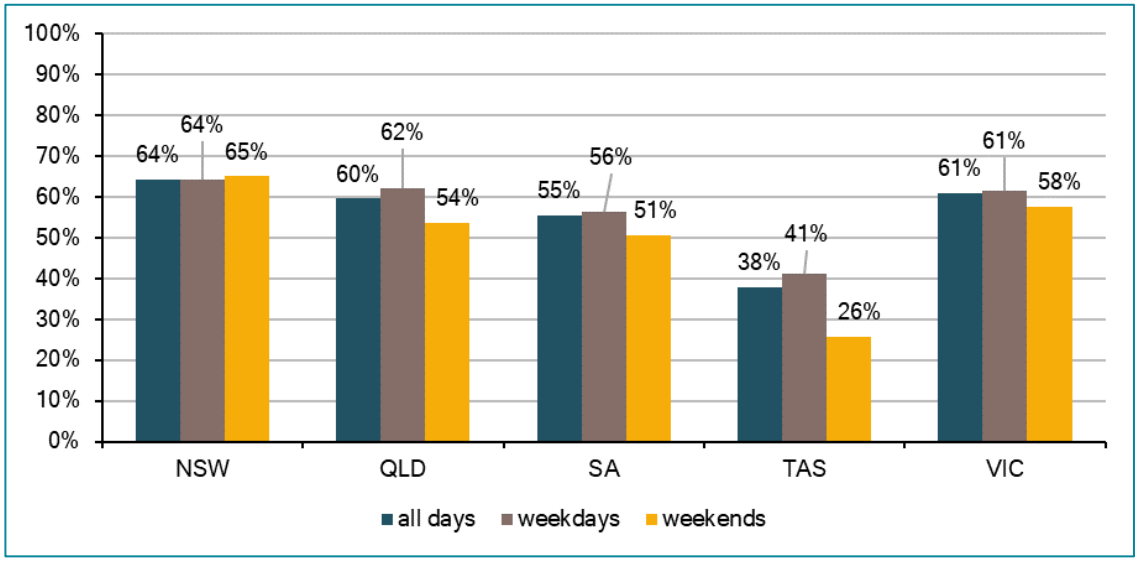


Figure 8: Number of high price trading intervals by trading interval in Q1 2019, by NEM region

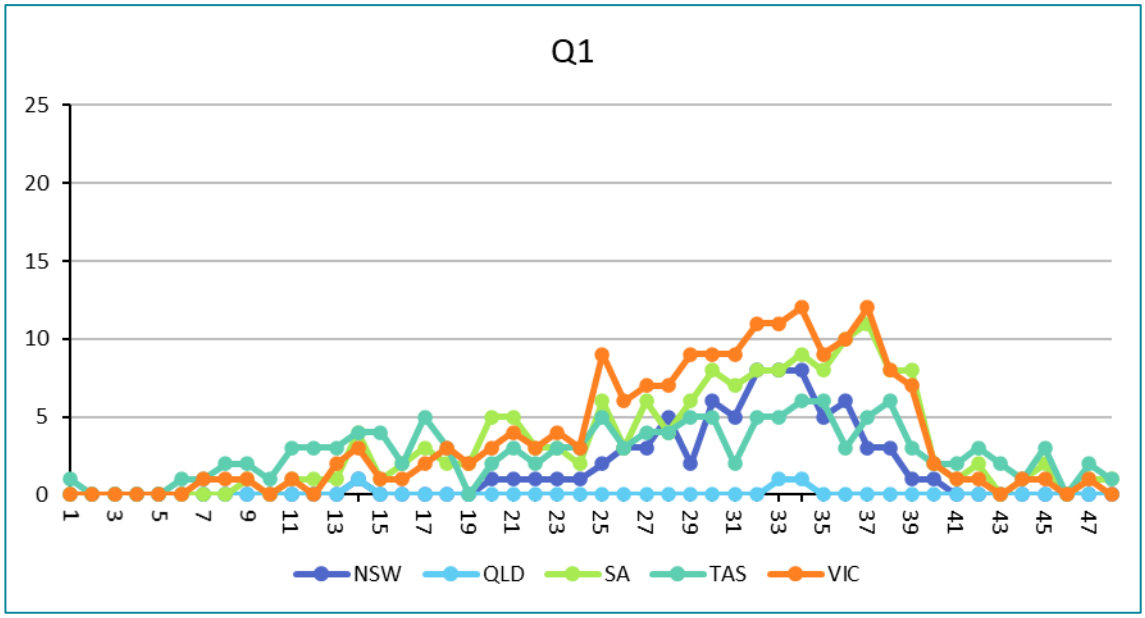


Figure 9: Number of high price trading intervals by trading interval in Q2 2019, by NEM region

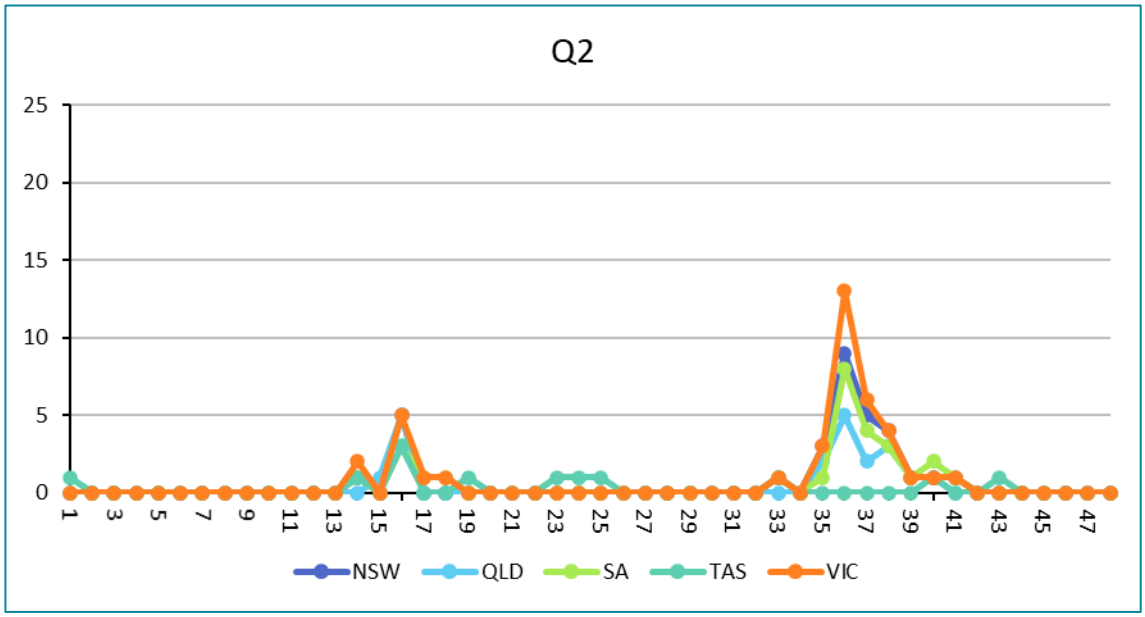


Figure 10: Number of high price trading intervals by trading interval in Q3 2019, by NEM region

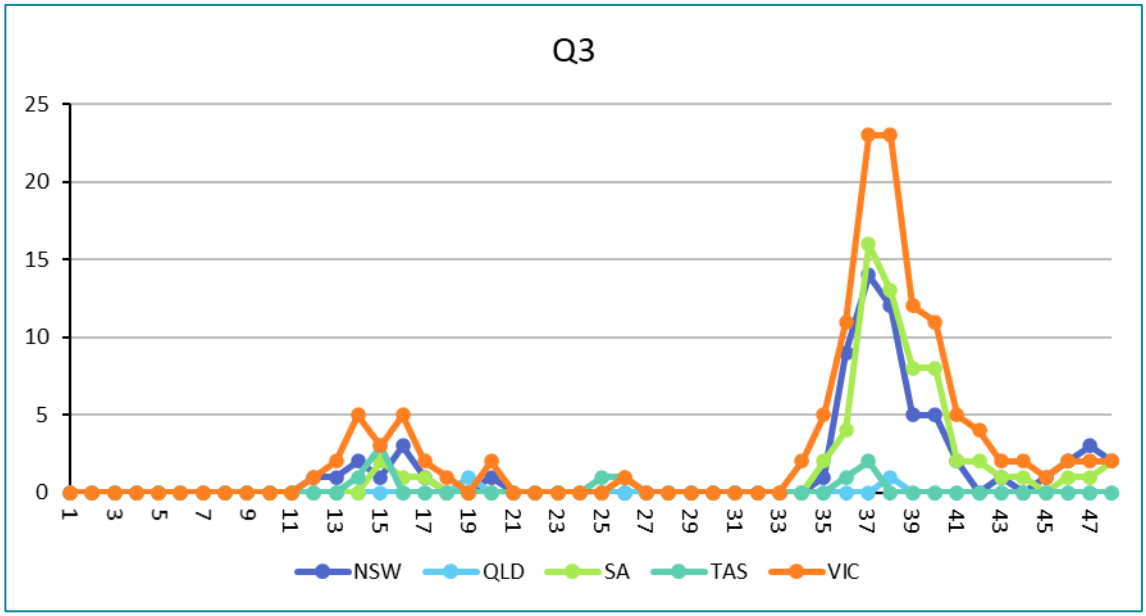
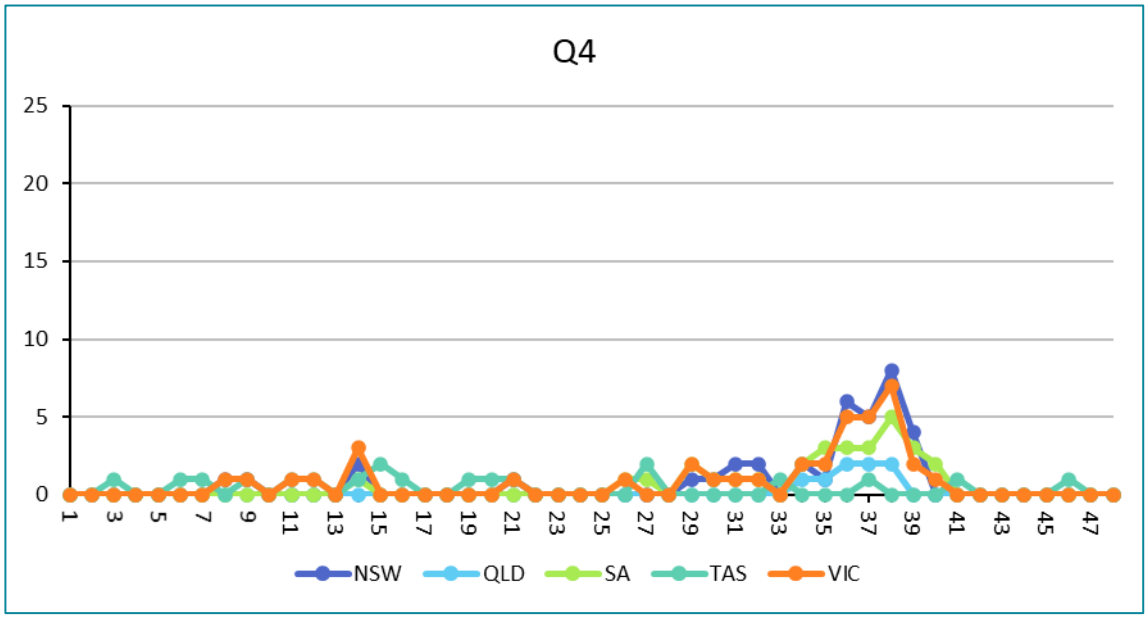


Figure 11: Number of high price trading intervals by trading interval in Q4 2019, by NEM region



Appendix D: Additional participation tables

Additional tables showing the proportion of NMIs that would be eligible in each region in each quarter of 2019 based on each of the three threshold levels for each segment are included in embedded spreadsheets that can be accessed through the link below.

For the data spreadsheet click the graph icon >>>>>