Attachment 13.1

Analytics + Data Science Report on Methodology for setting the service standard benchmarks and targets

Revised proposed access arrangement information

14 June 2018

Access arrangement information for the period
1 July 2017 to 30 June 2022

westernpower

Methodology for setting the service standard benchmarks and targets - Expert Report

Prepared for Western Power

6 June 2018

Analytics +
Data Science

Table of contents

List of figures

List of tables

1. Introduction

1.1. Scope of review

Analytics + Data Science (a+ds) was invited by Western Power to undertake an expert evaluation of the methodology for determining Service Standard Benchmarks (SSBs) and Service Standard Targets (SSTs) that will be adopted for Western Power's fourth access arrangement (AA4).

The terms of reference required a+ds to consider:

- Western Power's proposed methodology for setting SSBs and SSTs, as set out in the document *Fitting Distributions for AA4 Service Standard KPIs-Setting the Service Standard Benchmark (SSB) and Service Standard Target (SST);*
- the requirements of Sections 11.1, 5.6, 6.26, and 6.29 to 6.31 of the *Electricity Networks Access Code 2004* (the Code), and whether Western Power's proposed methodology is consistent with these requirements; and
- issues raised by the Economic Regulation Authority (ERA) in the *Draft Decision on Proposed Revisions to the Access Arrangement for the Western Power Network* (the Draft Decision) in relation to Western Power's proposed methodology, with specific reference to a) the setting of benchmarks at the 99th percentile level; and b) the process of averaging of percentile values from candidate probability distributions.

This document was developed in compliance with the Expert Witness Guidelines as set out by the Federal Court of Australia's Expert Evidence Practice Note GPN-EXPT. The relevant declarations can be found in Appendix A.

1.2. Key findings

The findings of a+ds' evaluation are as follows:

- Western Power's use of an averaging methodology for calculating SSBs and SSTs is consistent with multimodel averaging techniques widely employed in the field of inferential statistics;
- the use of a one per cent threshold to determine which candidate distributions are selected for inclusion in the averaging process is conceptually consistent with methods set out in peer reviewed literature. Furthermore, a+ds has not identified a consensus view that one methodology is superior to the alternative used by Western Power;
- a 99th percentile level for setting SSBs is consistent with the requirements of the Code in providing an incentive to maintain service standards;
- the ERA's proposed application of a 97.5th percentile level for setting SSBs does not guarantee that service standards will improve for those subset of customers experiencing below-standard service; and
- the information provided by Western Power in their technical report *Fitting Distributions for AA4 Service Standard KPIs-Setting the Service Standard Benchmark (SSB) and Service Standard Target (SST)* is sufficiently detailed to enable a user (or potential user) of the reference service to make a detailed assessment, when considering the reference tariff, of the economic costs, benefits, and risks of using said reference service.

2. Averaging methodology to generate robust SSB/SST estimates

2.1. Summary of Western Power's methodology

For each combination of SSB/SST measure and statistical distribution, Western Power has:

1. fitted five years of rolling average data using maximum likelihood estimation onto a selection of statistical distributions to generate candidate models;
2. performed a visual inspection of the raw data against the fitted distribution using quartile-quantile (Q-Q) and percentile-percentile (P-P) plots;
3. determined the theoretical distributions' goodness-of-fit using the Anderson-Darling test (Anderson & Darling, 1954);
4. discarded any candidate models from further evaluation where the p-value from the Anderson-Darling test is below a threshold value of 0.05;
5. calculated the relative quality of the remaining candidate models via the Akaike Information Criterion (AIC) and ranking the distributions according to their AIC in descending order;
6. discarded any candidate models where the difference between the AIC value of the candidate model and the AIC value of the candidate model with the lowest AIC exceeds one per cent;
7. calculated the 1%, 2.5%, 50%, 97.5% and 99% quantiles from the theoretical distribution with the lowest AIC value; and
8. calculated the average percentile values across the candidate distributions.

If all statistical distributions have an Anderson-Darling p-value of under 0.05 then quantiles were sampled directly from the historical data.

Concerns raised by the ERA

In the Draft Decision, the ERA has noted its concerns relating to the use of an averaging process across multiple candidate distributions (step 8) and the use of a threshold AIC value to restrict the number of candidate distributions used in the averaging process (step 6). These concerns are discussed in more detail below.

2.2. Multimodel averaging of percentile values

Objectives of the distribution fitting methodology

The purpose of fitting statistical distributions to observed performance data is to model the underlying physical processes that impact Western Power's service standards. Even under ideal circumstances, the complex dynamics that drive real world physical processes can't be perfectly captured within a finite sample of data. Therefore, any model of the underlying physical processes is, at best, only an approximation[1].

The objective underpinning the distribution fitting process is not to attempt to identify the one single "true" model. Instead, the methodology should accurately model, as far as is practicable, those physical processes that drive SSB/SST variability in a consistent and replicable manner. For this reason the use of a multimodel averaging process is desirable if it results in a more robust statistical model of the physical processes that impact service standards than can be obtained from any one single statistical model.

Model averaging as a response to data instability

The specification of one particular model as the "best" model for determining SSB/SST percentile values is inconsistent with the standard approach used by peer reviewed studies into statistical inference. In a practical manner, the "best" model is likely to vary from data set to data set, even if replicate data is captured from the same underlying process (Burnham & Anderson, 2002, p.151). The effect is not limited only to problems with small sample sizes. With data sets of even a moderate size, a slight change in the data may lead to the selection of a different model (Zou & Yang, 2004, p.70).

Burnham & Anderson (2004) provide a more complete discussion of the conclusions from multiple studies that demonstrate that a multimodel averaging approach is superior to the methodology in which parameter estimates are obtained from only the single "best" model.

Consequently, a+ds considers that Western Power's decision to base SSB/SST percentile values on a multimodel approach is consistent with the state-of-the-art practice in statistical inference.

[1] For example, see Burnham & Anderson (2004, p.264)

2.3. Selection of candidate models

Western Power's approach

Western Power's approach is to select candidate models where the AIC value of each model is within one per cent of the model with the lowest AIC. The mean percentile value of the selected candidate models is then calculated, and the process repeated for each SSB/SST.

The use of the AIC as a means to construct a relative ranking of alternative statistical models is common practice. For example, see Symonds & Moussalli (2011).

Selection of candidate models

That multiple statistical models are used in a multimodel framework does not imply that all available statistical models should be included. The rejection of some statistical models is common practice (Symonds & Moussalli, 2011, p.17), as the relative utility of each additional statistical model diminishes.

A number of alternative approaches have been developed to filter and weight the relative contribution of statistical models. Two such alternatives are discussed in detail in Burham & Anderson (2002), and Symonds & Moussalli (2011).

The simplest approaches involve rejecting any statistical model with an AIC value that differs more than a threshold value from the lowest observed AIC. In this approach:

$$\Delta_i = AIC_i - AIC_{min}$$

Such that any model (i) where Δ_i is greater than the threshold is rejected. Burnham & Anderson (2002) propose threshold values including 2 and 6. Alternatively, Richards (2005) proposes that any Δ_i less than 10 be considered an acceptable model.

Critically, there is no single threshold value for which there is uniform agreement across all authors.

A more complex methodology is set out in Symonds & Moussalli (2011) in which a weight is calculated from each Δ_i value. The weight adjusts each model's contribution to the aggregate multimodel. A cumulative sum of the weights is then calculated, and any model with the cumulative sum below a threshold value be considered for evaluation.

The weight is calculated as:

$$w_i = \frac{exp\left(-\frac{1}{2}\Delta_i\right)}{\sum exp\left(-\frac{1}{2}\Delta_r\right)}$$

The recommended threshold value differs between authors. Burnham & Anderson (2002) propose a threshold of 0.95, while Symonds & Moussalli (2011) recommend 0.90. The lack of consensus does not imply that any particular value is arbitrary, only that some discretion remains with the analyst.

Table 1 below illustrates an example of the AIC weighting methodology. In the example, the final distribution (gamma) would be rejected assuming a threshold value of 0.95. The weighting methodology delivers, in this case, a selection of candidate models that is the same as if Western Power's methodology were applied. Evidently, small changes in the assumed threshold value will impact the final multimodel output.

Table 1: Example selection process using the AIC weighting approach

Distribution	AIC	Δ_i	w_i	Cumulative w_i	$\frac{\Delta_i}{AIC_{min}} - 1$
Weibull	415.48	0	0.43	0.43	0.00%
GEV	416.66	1.18	0.24	0.66	0.28%
Weibull3	417.26	1.78	0.18	0.84	0.43%
Normal	418.46	2.98	0.10	0.94	0.72%
Gamma	419.64	4.16	0.05	0.99	1.01%

Western Power's methodology for selecting candidate models differs in some respects from those outlined above. However, Western Power's approach also follows the general pattern of using AIC values to establish a ranking of candidate models, and a threshold based on the distance from the lowest observed AIC value.

All methods depend on the use of a threshold value, for which there is no uniquely agreed value in the peer-reviewed literature.

Consequently, Western Power's methodology for selecting candidate models on which SSB/SST values are based is consistent with best practice approaches set out in peer-reviewed literature, noting that alternative but complementary approaches exist.

2.4. Variability in the composition of candidate models selected for averaging

In paragraph 1018(2) of the Draft Decision, the ERA notes that "the composition and number of distributions selected within the threshold value are likely to vary with time, introducing volatility and uncertainty". The ERA's observations are valid. The selection of candidate statistical models may change over time when using Western Power's methodology.

However, the alternative solution of selecting a single statistical model will only serve to exacerbate this source of variability. A change in the composition of which models are selected will have less of an effect on the percentile estimates than shifting entirely from one single distribution to another single distribution. If intertemporal consistency is indeed a priority, then the preference should be for Western Power's averaging methodology over the selection of a single distribution.

3. Use of the 99th percentile threshold for setting SSBs

3.1. Summary of Western Power's methodology

Western Power has proposed to use the 99th percentile values of the fitted distributions to set all SSB values in AA4, with the exception of call centre performance which is set at the 1st percentile.

SSBs were set at the 97.5th in AA3.

As part of the Draft Decision, the ERA rejected Western Power's proposed use of the 99th percentile value, instead proposing the use of the 97.5th percentile, consistent with AA3. Paragraph 1032 of the Draft Decision notes that the alignment of service standards with customer expectations was the most relevant factor in the ERA's decision.

3.2. Lower SSBs do not guarantee improved services for all customers

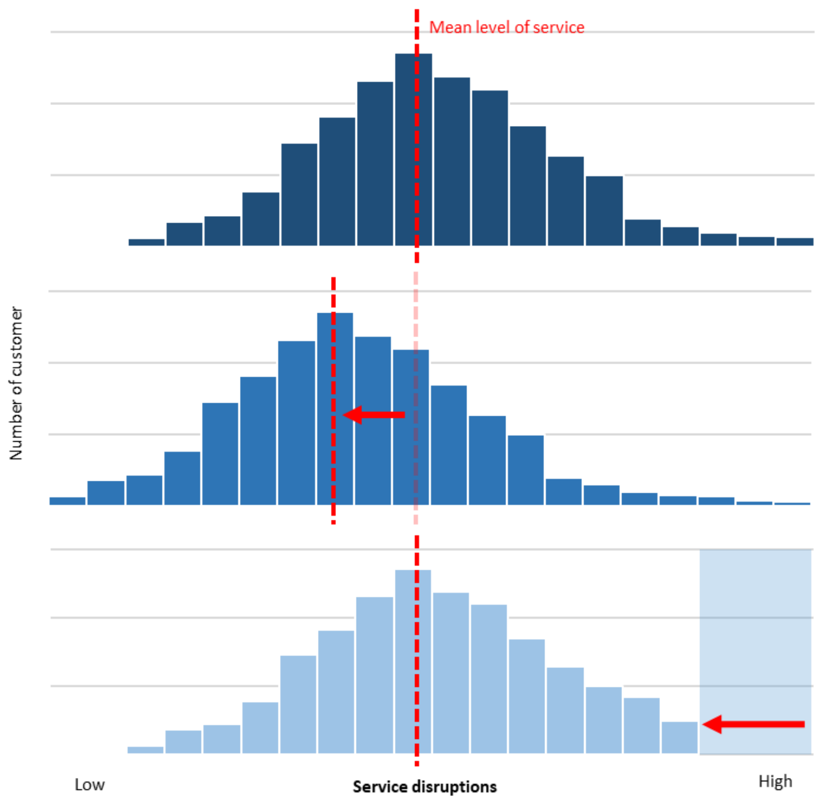
In Paragraph 1042 of the Draft Decision, the ERA notes that "customers in general have expressed satisfaction with current service levels", however that "a small proportion of customers may be consistently receiving below-standard service". It is on the basis of the experience of a small proportion of customers that the ERA has rejected Western Power's proposed use of the 99th percentile value in favour of a more stringent threshold.

Critically, there is no mention in the Draft Decision that the ERA has identified a need for an improvement in the *average* level of service provided to customers within each benchmark. The ERA has, however, noted a preference that the level of *variability* in service standards experienced between customers be decreased.

Modifying the SSBs to use the 97.5th percentile does not guarantee an improvement in service standards to each and every customer. It is not a targeted incentive mechanism. While service standards will improve for all customers *on average*, those small proportion of customers identified by the ERA as receiving below-standard service may not observe any change in performance standards with the ERA's proposed 97.5th percentile SSB.

Figure 1 provides an example of the difference between an increase in the average level of service (illustrated as a reduction in disruptions) and an improvement for only those customers experiencing below-standard levels of service. Figure 1A illustrates a hypothetical distribution of the level of service experienced by customers during AA3. Using the 97.5th percentile during AA4 will decrease the average number of disruptions during AA4, as shown in Figure 1B. Note that the right hand tail of the distribution – those customers experiencing below-standard service - has not changed. In contrast, Figure 1C illustrates the desired outcome of the ERA, as noted in the Draft Decision. The average level of service received by customers does not change, with any improvement limited to those customers experiencing higher levels of service disruptions.

Figure 1: Improvements in performance standards – reduction in mean or truncation?



If the intention of the ERA is to improve service standards for Western Power's customers on average, then the adoption of the 97.5th percentile would be an effective mechanism to incentivise such behaviour. However, if the intention of the ERA is to maintain average service standards at their current levels while improving standards to a targeted group of customers, then an alternative, appropriate regulatory mechanism should be identified to avoid incentivising inefficient investment behaviours contrary to the objectives of the Access Code.

In summary, Western Power and the ERA concur that the average level of service experienced by customers is consistent with customer expectations. Western Power noted that this level of service will be maintained by adopting the 99th percentile value for SSBs, consistent with the requirements of Section 11.1 of the Access Code. The ERA noted a preference for an improvement in the level of service to specific customers (not customers on average), and propose to achieve this outcome with a 97.5th percentile SSB. However, in the view of a+ds, adoption of the 97.5th percentile SSB will not necessarily ensure an improvement in service standards for those targeted customers, given it is an aggregate rather than targeted mechanism. For these reasons we see no rationale for not adopting the 99th percentile SSB value consistent with Western Power's proposed approach.

4. Provision of detailed information to users

4.1. Requirements to provide information to users

Section 5.6(b) of the Code requires Western Power to provide information that is "sufficiently detailed and complete to enable a user or applicant to determine the value represented by the reference service at the reference tariff". With reference to SSBs, a+ds interprets Section 5.6(b) to require Western Power to provide information at a sufficient level of detail such that a reference service user (or potential user) can quantitatively assess the risk of supply interruption they will face when using a reference service.

We consider that the value of the reference service is a function of any private benefit to the customer derived from accessing the service, the cost of using the service (the reference tariff), and the risk of supply disruptions. Interactions between these three variables occur. A higher probability of supply disruptions may decrease the expected private benefit derived from the service, or increase expected costs. Quantifying risk exposure is therefore critical for the user (or potential user) to accurately estimate the value represented by the reference service.

4.2. Information provided by Western Power

Western Power's report *Fitting Distributions for AA4 Service Standard KPIs-Setting the Service Standard Benchmark (SSB) and Service Standard Target (SST)* provides a detailed explanation of the service standard setting process. Information contained in the report includes:

- historic data in the form of charts;
- a detailed step-by-step overview of the methodology used to determine SSBs and SSTs from the raw data; and
- the technical description and five percentile estimates for each fitted probability distribution, enabling the replication of these distributions in an appropriate level of detail to undertake simulation modelling.

In our opinion, a reference service user (or potential user) could readily apply the information provided by Western Power to assess the quantity of supply disruptions occurring, and the duration of each disruption during AA4. That information is sufficient to enable a customer to develop an appropriately robust probabilistic estimate of the value of the reference service.

It is relevant to note that any assessment would not be an exercise in forecasting future disruption events, but rather an probabilistic assessment of the risk exposure faced by the reference service user (or potential user).

As defined by the Code, the service standard is an average value for a group of customers. A user (or potential user) may see value in obtaining a more granular assessment of their risk exposure based on their unique circumstances. However, it is questionable as to whether any such additional information would yield more informative results relative to the information already provided, or ultimately change the estimate of value for the user (or potential user). As an average value for a group of users, the service standard information provided by Western Power is a robust estimate of the service levels a user (or potential user) will receive. With this information a service user (or potential user) can evaluate their current or planned level of service against the benchmark, and develop a reliable estimate of service value.

For the reasons set out above, it is the opinion of a+ds that Western Power has met its obligations contained in Section 5.6 of the Code to provide sufficiently detailed and complete information such that a reference service user (or potential user) can quantitatively assess the risk of supply interruption they will face when using a reference service.

References

Anderson, T.W. and Darling, D.A. (1954). "A Test of Goodness-of-Fit". *Journal of the American Statistical Association*. Vol. 49. 765–769.

Burnham, K.P. & Anderson, D.R. (2002). *Model Selection and Multimodel inference: A Practical Information-Theoretic Approach (2nd Edition)*. Springer Publishing.

Burnham, K.P. & Anderson, D.R. (2004). "Multimodel Inference: Understanding AIC and BIC in model selection". *Sociological Methods and Research*. Vol 33(2). 261-304.

Lean, Y.U., Shouyang, W., Lai, K.K., & Nakamori, Y. (2005). "Time series forecasting with multiple candidate models: selecting or combining?" *Journal of Systems Science and Complexity*. Vol 18(1). 1-18.

Moussalli, A. & Symonds, M.R.E. (2011). "A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion". *Behavioural Ecology and Sociobiology*. Vol 65. 13-21.

Richards, S.A. (2005). "Testing ecological theory using information-theoretic approaches: examples and cautionary results". *Ecology*. Vol 86. 2804-2814.

Yang, Y. (2003). "Regression with multiple candidate models: selecting or mixing". *Statistica Sinica*. 783-908.

Zou, H. & Yang, Y. (2004). "Combining time series models for forecasting". *International Journal of Forecasting*. Vol 20. 69-84.

Appendix A: Declaration

This report has been prepared by Peter Shardlow of Analytics and Data Science Australia Pty Ltd (ACN 614 312 346).

As the author of this report I have read, understood and complied with the Expert Witness Guidelines entitled Expert Witnesses in Proceedings in the Federal Court of Australia (as defined in the Federal Court of Australia's Expert Evidence Practice Note GPN-EXPT). As the author I have made all the inquiries that I believe are desirable and appropriate and that no matters of significance that I regard as relevant have, to my knowledge, been withheld from this report.

A curriculum vitae for Peter Shardlow has been provided as Appendix B.

Appendix B: Curriculum Vitae

Peter Shardlow assists organisations to make better decisions through the application of quantitative analysis, with an emphasis on the fields of statistics, artificial intelligence, and machine learning.

Peter has over ten years' experience in identifying opportunities for organisations to employ quantitative analysis to leverage their unique competitive advantage, reduce costs and mitigate uncertainty. His expertise enables clients to make sound investment decisions by identifying the benefits and risks of employing solutions ranging from traditional statistical models to artificial intelligence-based technologies.

Qualifications

- University of Western Australia - Bachelor of Economics (Honours)
- University of Western Australia - Bachelor of Science

Experience

Date	Role
2016-2018	Principal Data Scientist, Analytics + Data Science Australia
2015-2016	Technical Specialist, Operations and Technology (WA), Australian Energy Market Operator
2013-2015	Supply Chain Analyst, Operations Research, BHP Billiton
2010-2013	Senior Consultant, Marsden Jacob Associates
2007-2010	Senior Analyst, Department of Treasury and Finance (WA)
2011	Lecturer, Public Economics, Curtin University of Technology
2010-2011	Sessional Instructor, Public Economics 312, Curtin University of Technology

Current Professional Affiliations

- Economics Society of Western Australia – Committee Member/Assistant Treasurer
- Institute of Analytics Professionals of Australia – Member.

Recent Projects

- Peak demand forecasting for the South West Interconnected System (2016-2018): Developed a predictive algorithm for identifying high electricity demand events, enabling retailers to reduce their customer's exposure to the Western Australian capacity market.
- Dispatch modelling in the Wholesale Electricity Market (2017): Undertook a peer review of electricity market dispatch modelling software developed by a regulatory agency in Western Australia.
- Application of machine learning techniques to enhance predictive maintenance capabilities for an Australian resources company (2016-2017): Design and implementation of analytical software to predict failure events in critical assets based on real-time equipment data.
- Geospatial analysis of electricity demand patterns (2017): Using geospatial statistical techniques, provided recommendations to a new energy business on the optimal location for a pilot program of their technology based on consumer demand patterns, rooftop solar PV generation levels, and social factors.
- Statistical analysis of electricity market dispatch dynamics in the WEM (2018): Using statistical and machine learning techniques, developed a robust algorithm for predicting bidding behaviours in the Western Australian electricity market.