Attachment 6.1

Review of service standards methodology

Access Arrangement Information

2 October 2017

Access Arrangement Information (AAI) for the period
1 July 2017 to 30 June 2022

westernpower

Review of service standards methodology

A report prepared for Western Power

18 September 2017

Analytics +
Data Science

Contact Information

PO Box 7178
Cloisters Square WA 6850
contact@analyticsdatascience.com
http://www.analyticsdatascience.com
ABN: 24 614 312 346

Table of contents

List of figures

List of tables

1. Introduction

1.1. Overview

Analytics + Data Science (a+ds) was engaged by Western Power to review the methodology adopted by the Insights and Analytics team for setting service standard key performance indicators as part of Access Arrangement 4 (AA4). Western Power's methodology centres on the fitting of statistical distributions to historic performance metrics in order to define Service Standard Benchmarks (SSBs) and Service Standard Targets (SSTs) for the AA4 regulatory period.

The scope of a+ds' engagement was defined as:

- undertaking an evaluation of the statistical methodology used to set each key performance indicator including distribution fitting;
- undertaking a technical review of the software code developed by Western Power for the purpose of estimating each key performance indicator; and
- providing an opinion on whether the methodology is fit for purpose for the setting of service standard benchmark targets to apply during the AA4 period.

1.2. Report structure

This report is structured around the following sections.

- Section 2 details the source material provided to a+ds for the review and any associated assumptions that underpin the review process.
- Section 3 provides an evaluation of the statistical methodology used by Western Power to define SSBs and SSTs.
- Section 4 discusses the implementation of the distribution fitting methodology in software code.

1.3. Key findings and recommendations

The key findings arising from our evaluation of the distribution fitting methodology and Western Power's response to the recommended actions are discussed in the summary table below.

Table 1: Summary of key findings and recommendations

Item	Finding
1	The outputs of Western Power's SSB and SST calculations were successfully replicated by a+ds using the source code and data provided.
2	We are able to confirm that the methodology as implemented by Western Power in their source code aligns with the methodology set out in their documentation.
3	In the view of a+ds, the methodology as implemented represents an appropriate methodology for setting SSBs and SSTs using a theoretically consistent and industry standard approach.
4	The data was evaluated against an appropriate selection of statistical distributions. We are not aware of any statistically-valid rationale for including alternative distributions in Western Power's calculations.
5	The application of a five-year regulatory period is an appropriate period for determining SSB/SST values. The use of shorter or longer periods of time which do not align with the duration of the regulatory period is not recommended.
6	The use of an averaging approach to select appropriate distributions overcomes an instability problem observed when fitting distributions to Western Power's data. Western Power's approach provides certainty to all parties that the proposed SSB/SST will not vary significantly over time.
7	The use of distribution's 99^{th} percentile value for setting SSB/SST aligns with Western Power's intention to maintain service standards over the AA4 period.

2. Source Material and Assumptions

2.1. Source material provided by Western Power

The source materials provided to a+ds consistent of reports, R source code and spreadsheets. The spreadsheets contained input data and model outputs as generated by Western Power. The source materials were initially provided to a+ds by Western Power on 28 August 2017, and revised source materials were subsequently provided on 12 September 2017.

We understand that the internal Western Power report *Fitting Distributions for Reliability KPIs-Setting the Service Standard Benchmark and Service Standard Targets* and the associated source code was amended to reflect changes in response to a+ds' draft report and other issues identified by Western Power.

Any subsequent revisions by Western Power to the methodology, source code or data occurring after 12 September 2017 will not have been evaluated by a+ds.

2.2. Assumptions and review limitations

The following section sets out all assumptions and limitations pertaining to our review of the Western Power methodology.

Validation of data against source systems

The time series data provided by Western Power is assumed to be accurate, free from errors and complete. The spreadsheet provided by Western Power containing the input data specifically notes that the data represents a 12-month rolling averages.

Our scope of work did not extend to validating data providence and integrity from source systems.

Evaluation of data capture process

We have not evaluated the methodology by which service standards data is collected, filtered and aggregated. It is therefore not possible to determine whether input data was censored or truncated in any way during the data capture process as a consequence of physical or system-level limitations[1]. Censoring and truncation of data will influence the distribution fitting methodology.

Validation of source code not developed by Western Power

The R statistical software[2] used by Western Power, and in particular the source code developed for calculating SSBs/SSTs, makes use of a number of additional libraries beyond what are available in the base install. The accuracy of any software not developed by Western Power has only been validated insofar as the software produces repeatable and replicable results.

Further changes to input data prior to AA4 submission

We understand that Western Power intends to update the SSB/SST input data closer to the submission date of AA4 as new data becomes available to the business or as other data quality issues are identified. Changes to the input data may result in a misalignment between the estimates reviewed by a+ds and those submitted to the ERA.

[1] Beyond the zero lower bound.
[2] R Core Team (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at https://www.R-project.org/

3. Evaluation of distribution fitting methodology

3.1. Overview

The following section sets out a+ds' review of Western Power's SSB and SST distribution fitting methodology. The review considers whether:

- the distribution fitting methodology for SSBs and SSTs as documented by Western Power represent an appropriate statistical methodology; and
- Western Power's software implementation of the distribution fitting methodology aligns with their description of the methodology as set out in the documentation.

In the view of a+ds, the methodology as outlined in this section represents an appropriate methodology for the purpose of selecting a statistical distribution for setting SSBs and SSTs using a theoretically consistent and industry standard approach.

3.2. Western Power's methodology for fitting statistical distributions

Western Power's methodology for fitting statistical distributions to determine SSBs and SSTs is set out in the report *Fitting Distributions for Reliability KPIs-Setting the Service Standard Benchmark and Service Standard Target*. The report covers the procedures used to fit continuous and discrete distributions to specific service standard performance measures.

For the sake of completeness, a summary of Western Power's distribution fitting methodology is set out below. A list of the statistical distributions considered by Western Power is also included in Appendix A.

Methodology for fitting continuous distributions

Our understanding of Western Power's approach is described below. The description is based on the updated source code and documentation provided to a+ds on 12 September 2017.

For each combination of performance measure and statistical distribution, Western Power has:

- fitted the chosen statistical distribution onto five years of rolling average data using maximum likelihood estimation;

- performed a visual inspection of the raw data against the fitted distribution using quartile-quantile (Q-Q) and percentile-percentile (P-P) plots;
- determined the theoretical distributions' goodness-of-fit using the Anderson-Darling[3] test;
- discarded any distributions from further evaluation where the p-value from the Anderson-Darling test is below a threshold value of 0.05;
- calculated the relative quality of the remaining statistical models via the Akaike Information Criterion (AIC) and ranking the distributions according to their AIC in descending order; and
- calculated the 1%, 2.5%, 50%, 97.5% and 99% quantiles from the theoretical distribution with the lowest AIC value.

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

The SSB/SST target for each measure is then derived as the mean of the 99th percentile values from all distributions with an AIC value within one per cent of the lowest AIC value for that performance metric.

Methodology for fitting discrete distributions

For discrete distributions, Pearson's Chi-Squared test was applied in place of the Anderson-Darling test.

3.3. Choice of a five-year sample period

Western Power has adopted a five-year time duration for fitting statistical distributions to the performance data. We note that the five-year duration aligns with the length of the regulatory period, providing an interval in which financial and regulatory incentives remained consistent.

Setting the SSB/SST using a smaller sample of data would result in greater uncertainty over the accuracy of the distribution fit. This effect is manifested in terms of higher standard errors for the estimated shape parameters. For the purposes of providing a robust and reliable estimate, greater uncertainty is undesirable.

---

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

Time periods longer than five years will consider historic performance during an interval in which Western Power was responding to very different financial and regulatory incentives. If a longer time horizon were to be chosen, it would then be necessary to apply relevant adjustments to the time series data to take account of exogenous changes in regulatory and financial incentives. In this situation, we recommend that Western Power adopts the principle of parsimony by retaining the five-year duration, rather than choosing an arbitrarily long period and then attempting to control for the exogenous effects of capital expenditure and changes in incentive frameworks.

For the reasons noted above, we agree with Western Power's decision to adopt a five-year period for calculating SSB/SST metrics.

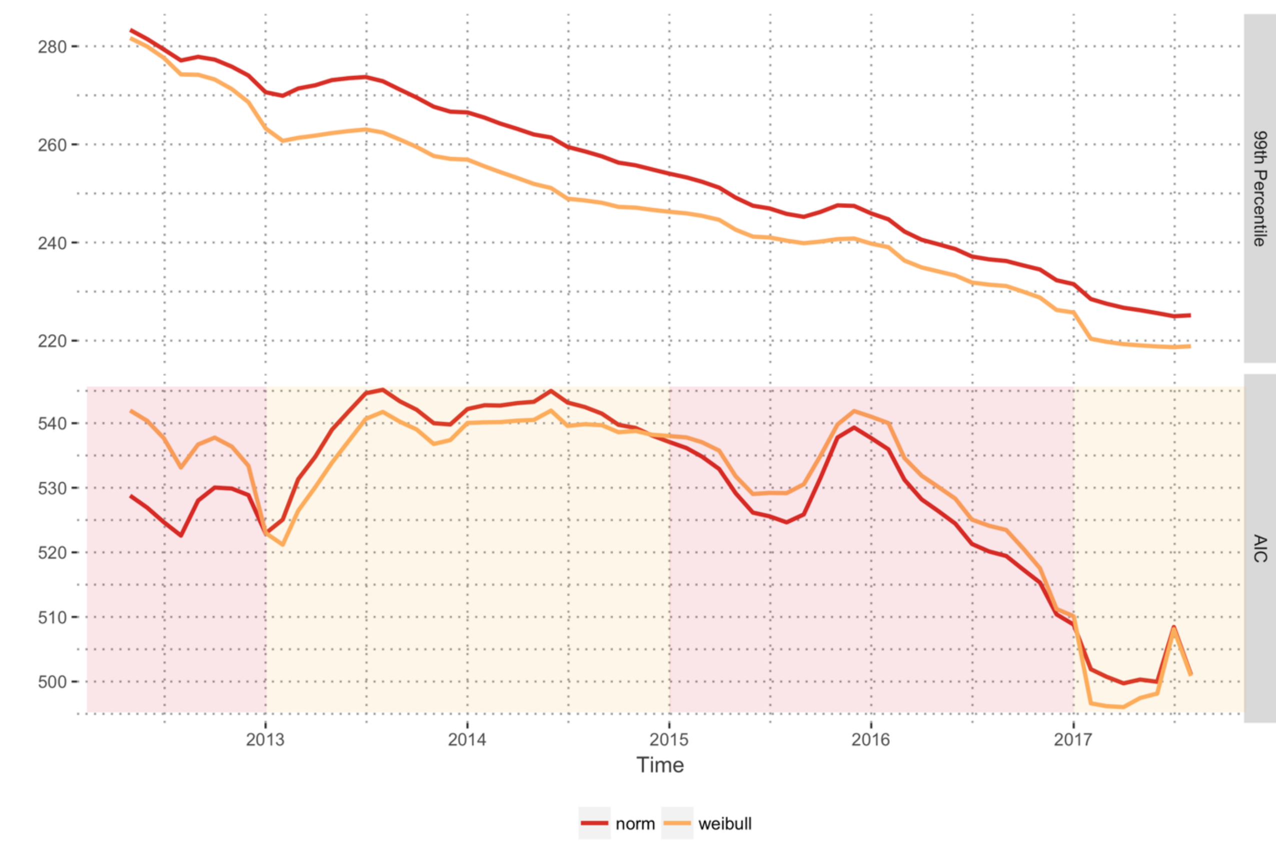
3.4. Averaging multiple distributions based on AIC threshold

Western Power uses the AIC estimate to determine which distributions most accurately reflect the underlying data generation process. The approach requires that the mean 99th percentile of the fitted distributions be calculated using those statistical distributions demonstrating an AIC value within one per cent of the lowest AIC value for each metric.

We understand that the purpose of averaging the 99th percentile values from multiple distributions is to overcome the instability associated with selecting a single distribution at a particular point in time based solely on the lowest AIC. Relying only on the distribution with the lowest AIC can result in substantial variance in the SSB/SST over time as the relative ordering of distributions (based on their relative AIC value) changes in response to new data. The instability is largely a function of different candidate distributions generating very similar AIC values.

This behaviour is demonstrated in Figure 1, which illustrates the AIC and 99th percentile value for Rural Short SAIDI using Normal and Weibull distributions on a rolling five-year basis. The figure is generated using data back to 2012. The x-axis represents the end-date of the five-year rolling time window.

Figure 1: AIC calculated on a 5-year rolling basis for Rural Short SAIDI



If Western Power were to rely only on the distribution with the lowest AIC value, then the preferred distribution, and the associated SSB/SST, would have changed four times. At each of these transition points the 'correct' 99th percentile to use would also have changed, resulting in jumps in the target of approximately 10 per cent on each occasion.

The instability created by relying solely on a single distribution is undesirable. It creates uncertainty for customers and Western Power on what level of performance should be achieved during AA4, at a time when Western Power is seeking only to maintain existing service standards in line with customer expectations.

One per cent AIC threshold

To determine SSBs/SSTs, Western Power calculates the mean 99th percentile value for all distributions with an AIC value within one per cent of the distribution with the lowest AIC value. The one per cent threshold ensures that the 99th percentile value for one or more distribution is included when deriving the SSB/SST. In a minority of cases only a single distribution was selected, on the basis that no other distribution demonstrated an AIC value within one per cent of the lowest value. Averaging a number of valid candidate distributions provides stability to the SSB/SST measures over time as the underlying data changes.

We note that alternative AIC cut-off thresholds could have been adopted. For example, of 0.5 per cent or 2.0 per cent. However, the objective of the averaging step is to provide stability to the SSB/SST measures by considering a small number of possible candidate distributions. A threshold that is too high would result in all distributions being considered, departing from the principle that the SSB/SST should be based on the distribution of best fit. A threshold that is too low would result in only the single best distribution being considered, negating the rationale for averaging the 99th percentile values.

The one per cent value provides a pragmatic trade-off between considering too many and too few candidate distributions. Based on the data available to a+ds, a lower threshold of 0.5 per cent would result in a number of SSB/SST calculations considering only a single distribution, while a 2.0 per cent threshold would result in those same calculations considering all possible candidate distributions. Keeping in mind the objective of this step in the calculation is only to stabilise the SSB/SST from 'jumping' between alternative distribution types over time, the one per cent threshold provides an appropriate balance given the AIC values as currently calculated.

Conclusion

For the reasons noted above we concur with Western Power's approach for averaging the 99th percentile values from candidate distributions and the use of a one per cent cut-off for determining which distributions to average.

3.5. Use of the 99th percentile of fitted distributions to determine SSB/SST

Western Power has used the 99th percentile values of the fitted distributions to set the SSB and SST metrics. We note that Western Power's objectives for the coming five-year period is to maintain current service standards performance in line with customer expectations. Given that objective, it is appropriate to choose a percentile value that does not penalise Western Power for not continuing to improve performance. Choosing a lower threshold value would increase the probability that, in the absence of further investment (at the expense of customers), service standards would not be met and Western Power would be financially penalised.

We are also not aware of any statistical basis which would suggest the 99th percentile value to be any more or less appropriate than an alternative threshold. Consequently, we concur that the 99th percentile value an appropriate threshold that aligns with Western Power's longer term strategic objectives for AA4.

4. Review of source code implementation

4.1. Replication of results

a+ds has performed a technical review of the software code developed by Western Power for the purpose of estimating each service standard. No fundamental flaws were identified in the code base resulting from:

- incorrect implementation of the software code relative to the methodology as outlined in the technical report; or
- incorrect implementation due to coding omissions or errors.

Using the source code provided, we were able to replicate the numerical outputs generated by Western Power and confirm that the results for each SSB/SST align with the expected value of that output.

4.2. Replication of the software environment

The R software environment and its extensive array of libraries is constantly updated. This presents a challenge for analytics professionals working in commercial environments where the need to ensure replicability of results is critical. Minor changes to the software between releases can have significant effects on calculation outputs. Changes may alter input assumptions, default parameters or the underlying methodology.

While software version control was not identified as an issue when replicating Western Power's results as part of this review, we recommend that a stricter approach to software and package management be adopted to avoid replication issues in future.

4.3. Data version control

The input data spreadsheet provided by Western Power is accurate at one point in time. Discussions with Western Power suggests that this data is likely to be revised, either through the addition of new data or changes to existing values. In a production environment and while operating on a single network, the risk of using outdated data can be mitigated by having a single-source-of-truth. Western Power appears to manage the input data in this manner.

Once the input file moves to uncontrolled or offline environments, ensuring different users are all referencing the same version becomes more challenging. To mitigate the risk of users unintentionally comparing results that are constructed from different data sources, we recommend that Western Power implements a system for version control that relies on a unique file checksum, which can be readily compared by two different users to ensure consistency of input data.

Appendix A: Statistical distributions

Table 2: Discrete and continuous statistical distributions considered by Western Power

Distribution Type	Type
Historical	Continuous
Weibull (3 parameter)	Continuous
Weibull	Continuous
Gamma (3 parameter)	Continuous
Generalised Extreme Value	Continuous
Lognormal3	Continuous
Normal	Continuous
Log Logistic (3 Parameter)	Continuous
Logistic	Continuous
Gamma	Continuous
Log Logistic	Continuous
Exponential	Continuous
Lognormal	Continuous
Poisson	Discrete
Negative Binomial	Discrete
Binomial	Discrete
Geometric	Discrete