Attachment 6.2

Fitting Distributions for AA5 Service Standard Benchmarks

Access Arrangement Information

1 February 2022

westernpower

EDM 58052877

An appropriate citation for this paper is:

Access Arrangement Information - Fitting Distributions for AA5 Service Standard Benchmarks

Western Power

363 Wellington Street
Perth WA 6000
GPO Box L921 Perth WA 6842

T: 13 10 87 | Fax: 08 9225 2660
TTY 1800 13 13 51 | TIS 13 14 50

Electricity Networks Corporation
ABN 18 540 492 861

enquiry@westernpower.com.au
westernpower.com.au

Enquiries about this report should be directed to:

AA5 Project Team

Email: AA5@westernpower.com.au

Contents

List of tables

List of figures

1. Summary and Purpose

The purpose of this report is to document the method and outcome for developing performance metrics for the fifth Access Arrangement (**AA5**).

The performance metrics take two forms:

- Service Standard Targets (**SSTs**) under the Service Standard Adjustment Mechanism (**SSAM**) that dictate financial incentives/disincentives
- Service Standard Benchmarks (**SSBs**) that dictate minimum service standards

1.1 Reason for benchmarking

Section 11.1 of the Electricity Networks Access Code 2004 states that:

A service provider must provide reference services at a service standard at least equivalent to the service standard benchmarks set out in the access arrangement and must provide non-reference services to a service standard at least equivalent to the service standard in the access contract.

By establishing a benchmark, Western Power's customers can assess the level of service performance that they can reasonably expect and that performance can be measured over time. Implicitly, there is a trade-off between the cost of providing reference services and the likely benefit in the form of achievable performance levels.

1.2 Benefits of benchmarking

Measuring service standard performance provides the basis on which to identify deterioration in service as well as help quantify opportunities for improvement. By establishing an objective yardstick, Western Power can assess its performance and take corrective action as required. Moreover, Western Power is well positioned to determine the appropriate level of expenditure to ensure service remains within the required tolerances.

1.3 Principles guiding the benchmarking process

The objective is establishing a reasonable expectation given the inherent random variation or 'noisiness' in the data. The approach taken in this report is to formally apply a statistical distribution, which explicitly recognises the observed level of randomness.

However, it is important to note that in conducting the statistical analysis, close attention is given to ensuring that the benchmarking process is consistent with the following principles:

- The resulting benchmarks are practical to administer.
- They are adaptive, allowing for changes in the operating environment that are outside Western Power's control.
- While adaptive, the benchmarks should be relatively stable and with minor movement over time being predictable.
- The performance measures should be meaningful for customers.

Western Power believes that adopting these principles helps ensure a systematic approach to managing service standards over time. In practice, this means that the textbook statistical procedures have been modified to better serve these principles.

The primary outcome is a performance measurement system that conveys clear signals indicating when remedial action is required.

2. Method

Key messages:

- Western Power has based its method firmly on statistics literature. This section provides a detailed description of the standard approach identified in the literature review.
- The standard approach has two parts: visual inspection; and calculation of goodness-of-fit statistics. The visual inspection process, based on P-P and Q-Q plots, serves to visually confirm the statistical results.
- The goodness-of-fit statistics chosen by Western Power for evaluating candidate statistical distributions are the Anderson-Darling statistic for testing continuous distributions; and the Chi-squared statistic for discrete distributions. These statistics are used to reject statistical distributions that are poor representations of the underlying Data Generating Process.
- The Akaike Information Criterion (**AIC**) is used to rank the remaining candidate statistical distributions in terms of how well each distribution fits the data.[1]
- The statistical distribution of best fit is defined as the one passing the goodness-of-fit test and having the lowest AIC score.

2.1 Introduction

A current process for setting SSBs was developed and approved by the Economic Regulation Authority (**ERA**) from the third access arrangement (**AA3**) onwards. The process involves fitting Probability Density Functions (PDF) to historic 12 month rolling average performance data. It is proposed that the AA5 method build on this approach.

Note: The process for setting service standard targets (SSTs) under the service standard adjustment mechanism (SSAM) financial incentive scheme, is based on the simple average of the financial year performance for the prior access arrangement period. This is consistent with the approach used by the AER for setting financial targets under the service.

2.2 Discussion

Probability distribution fitting describes the process of determining the PDF that best represents the historic data. The objective of probability distribution fitting is to determine which PDF most accurately predicts the observed frequency of observations (i.e. performance measurements).

Before explaining the process for identifying the best fitting PDF, it is necessary to discuss two broad types of distributions. Namely, PDFs that model:

- Continuous variables (e.g. age, weight, height etc) that can be measured at any interval along a number line; and
- Discrete variables that measure the probability of countable outcomes (e.g. coin flips, customer complaints etc).

Irrespective of the type of random variable being investigated, the general distribution fitting process involves the following four steps:

---

[1] The lower the AIC score, the better the fit of the distribution to the Data Generating Process.

1. Model each candidate PDF
2. Estimate PDF parameters
3. Reject poor fitting PDFs using goodness-of-fit statistical test
4. Select the best PDF based on the quality of fit statistic.

2.3 Fitting continuous PDFs

Two methods for fitting a distribution to continuous variables are:

- Visual inspection of diagnostic plots
- Maximum Likelihood Estimation (**MLE**) and selection based on statistical tests; in this case the p-value of the Anderson-Darling statistic and the Akaike Information Criterion.

Western Power has applied both the visual and the MLE method to determining the best fitting distribution for each performance indicator.

Note that in fitting PDFs, there is an assumption that the chosen PDF is an accurate predictive model of the underlying Data Generating Process. That is, the chosen PDF is deemed to provide a reasonable reflection of the process that causes measured observations to deviate from the centre of the distribution of observations. Particularly important issues are:

- The stability of the measures of central tendency of the observations, such as the mean, median and mode. For example, there is often a strong assumption of a stationary mean. Time series data can exhibit trends, which implies an unstable centre of the distribution of observations. A secondary issue is whether any movement in the mean is due to systematic or random influences.
- The strength of the central tendency, which is measured in terms of the average distance from the centre of the PDF. The higher the average distance, the less strength in central tendency. This implies that the measures of central tendency could be poor predictors of future observations.
- The degree of any asymmetry in the collection of observations. Asymmetry implies that the probability of observing values below the measured centre of the distribution is not the same as the probability of observing values above the measured centre.

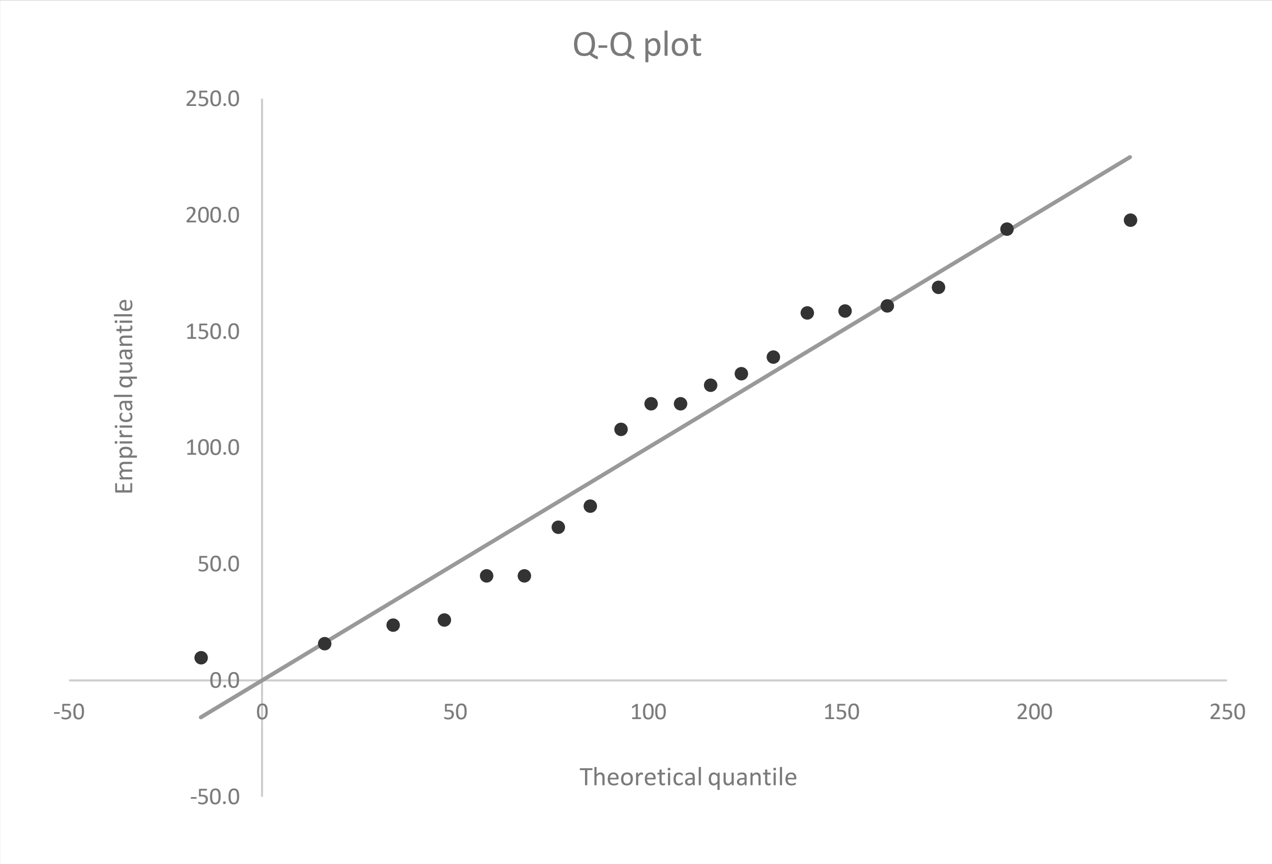
2.3.1 Visual inspection

The main visual diagnostic tool is the quantile-quantile (**Q-Q**) scatter plot of observations (i.e. data points) drawn from a specific PDF (e.g. the Normal distribution) against the sampled observations. Note that quantile refers to the division of a sample of observations into equal-sized subgroups. A plot showing quantiles provides a visual guide to the proportion of observations that are below a given value, for example half of observations lie below the median.

The Q-Q plot provides a visual guide of how well observations map to a chosen PDF. A Q-Q plot typically shows a scatter plot of observations against a reference line rising from left to right. A perfect match between observed values and theoretical values drawn from a chosen PDF would be indicated by all the plotted observations lying directly over the reference line.

When dealing with real world data, it is unlikely that all observations would lie perfectly along the reference line, even if the correct PDF is used as the benchmark in the Q-Q plot. This can make deciding which PDF is best to model differences within families of PDFs difficult. Choosing the best distribution is typically more difficult when there are few observations, e.g. less than 100.

Figure 1: An example of a Q-Q plot



When inspecting a Q-Q plot such as the one shown in Figure 1 , the main features to look for are:

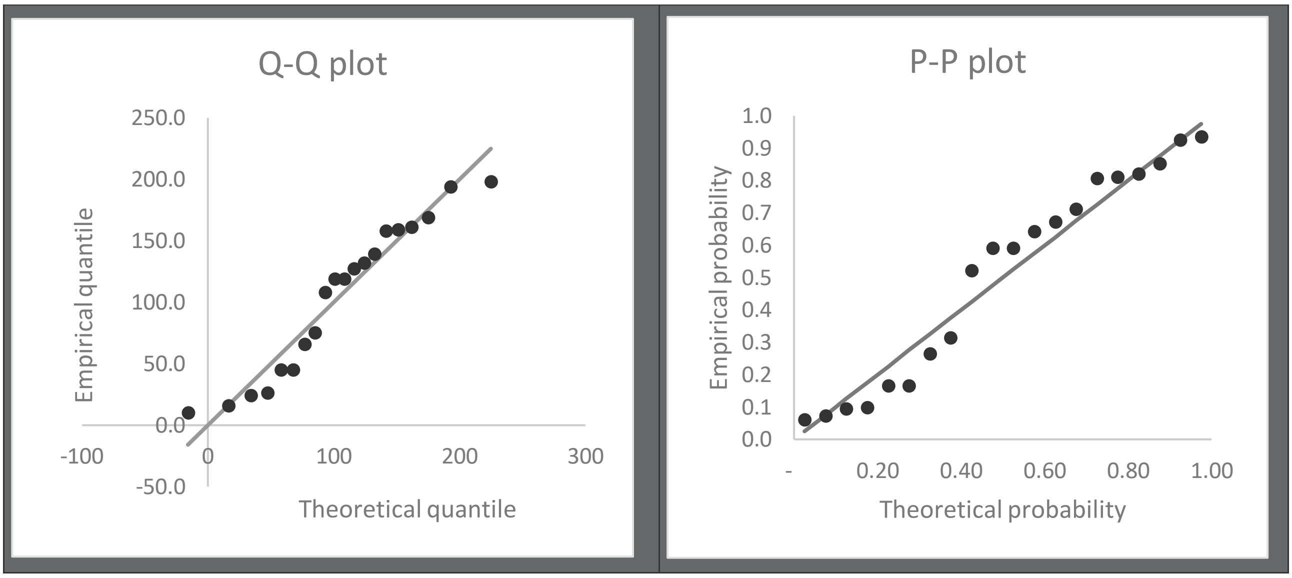
- The distance of low and high value observations from the reference line compared to the mid-value observations. In Figure 1, the observations at the low and high value ends are approximately the same distance from the reference line as the mid-value observations.
- Differences in the distance from the 45-degree line for low-value observations compared to high-value observations. In Figure 1, the two highest value observations lie further from the reference line than the lowest two value observations. This indicates asymmetry in the distribution of observations, but is insufficient to determine that the PDF of the Data Generating Process is asymmetric.
- Clustering of observations, which could indicate the most likely weighted centre of the distribution. There are no signs of clustering in Figure 1, which indicates an even spread of observations along the reference line.

The Q-Q plot is most informative when comparing plots with different theoretical PDFs. Typically, one of the PDFs will exhibit a better match to the observed values than the others. The best matching PDF is then chosen as the PDF to use in predictive modelling or benchmarking.

2.3.2 Q-Q plot versus P-P plot

In addition to the Q-Q plot, it can be helpful to also consider the probability-probability (**P-P**) plot. These plots are similar in intent, with the P-P plot comparing theoretical probabilities drawn from a benchmark PDF and the empirical probabilities calculated for each observation using the benchmark PDF adjusted for the sample mean and standard deviation.

Figure 2: An example of a Q-Q plot and a P-P plot for the same set of observations



The diagnostic information provided by Q-Q and P-P plots are slightly different. Figure 2 presents the Q-Q plot and the P-P plot for the same observations. Comparing the two, it is clearer that the Q-Q plot presents a comparison between the benchmark PDF and the observed data in the same scale as the original data. By contrast, the P-P plot compares the probability of the reference PDF.

The practical difference comes down to the Q-Q plot providing better visibility of the extremes of the distribution whereas the P-P plot provides better visibility of the centre of the distribution.

2.3.2.1 Concluding remarks on visual inspection

The visual inspection of the data using Q-Q and P-P plots were used to confirm that the statistics identified the best fitting plots.

2.3.3 Statistical measures – goodness of fit statistics

In addition to visual methods, there are statistical methods of determining the best fitting PDF. The three main statistics are the:

- Kolmogorov-Smirnov statistic
- Cramer-von Mises statistic
- Anderson-Darling statistic.

These test statistics are the mathematical equivalent of the visual tests.[2] That is, the test statistics compare the empirical distribution with a chosen theoretical distribution, determine the degree of difference and test whether the magnitude of the difference is large enough to reject the chosen theoretical distribution. That is, given $X_1, \cdots, X_n$ identically and independently drawn samples (i.e. observations) from an unknown distribution F, a formal hypothesis test can be established:

$$H_0 : F = F_0 \text{ vs. } H_1 : F \neq F_0$$

[2] This section draws on information provided by Rui Castro 2013: Lectures 2 and 3 – Goodness-of-Fit (GoF) Tests.

Where H_0 represents the null hypothesis, H_1 is the alternative hypothesis, F represents the empirical distribution and F_0 is the hypothesised distribution. According to Castro (2013), these tests are based on the Glivenko-Cantelli theorem, which states

$$\sup_t \left| \hat{F}_n(t) - F_o(t) \right| \xrightarrow{a.s} 0 \text{ as } n \to \infty$$

Where sup indicates the supremum (i.e. the least upper bound), $a.s.$ indicates asymptotic, $\hat{F}_n$is piece-wise constant and F_0 is ordered so that it is a non-decreasing function. In words, this theorem states that the supremum of the difference between the empirical and theoretical distributions asymptotically approaches zero as the sample size of observations approaches infinity.

This theorem and the tests that are based on it allow for some non-negative differences between the empirical and theoretical distributions. Hence, a threshold critical value needs to be determined which, if exceeded, indicates that the null hypothesis is rejected.

An important feature of all three tests is that the test statistics either do not depend at all on the theoretical distribution (e.g. the Kolmogorov-Smirnov test) or only partly on the theoretical distribution (i.e. the Anderson-Darling test) being tested against the sampled observations. In addition, these tests are consistent under any alternative hypothesis.

2.3.3.1 Kolmogorov-Smirnov statistic and test

The Kolmogorov-Smirnov test is calculated in a piece-wise method by:

- Arranging the observations in ascending order; and then
- Calculating the Empirical Cumulative Density Function (**ECDF**), which is a one-dimensional array containing a series of values derived by dividing the number of observations that are less than each observed value by the total number of observations; and then
- Calculating the maximum distance between the ECDF and the theoretical Cumulative Density Function of a chosen reference PDF.

The Kolmogorov-Smirnov statistic can be calculated for each PDF under consideration. The test is constructed as a hypothesis test:

- The null hypothesis is that the data is generated by the theoretical PDF being tested.
- The alternative hypothesis is that the data is not generated by the theoretical distribution being tested.

The null hypothesis is rejected if the test statistic exceeds a prescribed significance level.

2.3.3.2 Cramer-von Mises statistic and test

The Cramer-von Mises test is like the Kolmogorov-Smirnov test except it replaces the supremum with the product of the squared differences between the empirical and theoretical distributions.

2.3.3.3 Anderson-Darling statistic and test

The Anderson-Darling statistic is another variation of the Kolmogorov-Smirnov test where the difference between the empirical and theoretical distributions on the tails of the distribution receives more importance. Note that the critical values of the Anderson-Darling test are dependent on the theoretical distribution being tested.

According to the literature reviewed, the Anderson-Darling test has higher power than the other two tests. This means that it is better able to reject a false null hypothesis. For this reason, the goodness-of-fit results in this report focus exclusively on the Anderson-Darling test results. Specifically, the p-value of the Anderson-Darling is reported. To accept the candidate distribution as a good fit, the p-value needs to be greater than a critical threshold value.

2.3.3.4 Akaike Information Criterion

The Akaike Information Criterion (**AIC**) is a statistic that provides information about the relative quality of goodness-of-fit. A key difference between the AIC and the other goodness-of-fit statistics is that there is no null hypothesis to be rejected.

The statistic is calculated as follows:

$$AIC_c = 2k - 2ln(\hat{L})$$

Where k represents the number of model parameters and $ln(\hat{L})$ is the natural logarithm of the likelihood score. Note that the likelihood score is derived from maximising the likelihood function. So, there is a clear trade-off between adding more parameters and increasing the likelihood score. That is, for a more complex distribution to be accepted as best fitting, the likelihood score needs to increase (representing a benefit) by more than the cost of adding additional parameters.

This score guards against relying on unnecessarily complex models to determine the best fitting distribution.

2.3.3.5 Using the Anderson-Darling and AIC statistics together

By using both the Anderson-Darling and the AIC statistics, a clear rank can be formed among the candidate distributions for each service standard measure. This is achieved in two steps:

1. Fit each candidate distribution, calculate the Anderson-Darling p-value and reject all distributions registering a p-value less than 0.05.
2. Of those candidate distributions left, choose the distribution that registers the lowest AIC score.

2.3.3.6 Consideration of the underlying assumptions

Fitting statistical distributions requires adherence to the assumptions underpinning the validity of the chosen statistics. The most important assumptions are:

- The mean and variance of the underlying Data Generating Process are constant;
- That each observation is independent of the other observations contained in the sample data; and
- For a given service standard measure, that each observation is drawn from the same univariate distribution.

These are strong assumptions that could be inadvertently violated. In using real world data, generally accepted practice is to relax these assumptions somewhat. That is, the chosen distribution is regarded as the best approximation of the unknown Data Generating Process.

In some cases (such as a strongly trending mean) relaxing the underlying assumptions could result in gross errors. In such cases, it may be reasonable to adjust the distribution fitting process by:

- Segmenting the sample data and independently fitting the distributions to each segment.

- Pre-whitening the data.

Choosing the appropriate action depends on several factors:

- The overall sample size for each service standard measure – reducing sample size may reduce precision.
- Identifying any underlying causes of a non-stationary mean and/or variance.

Fitting distributions to subsets of the data can be informative. If dividing the data in different ways leads to radically different results, then it can be reasonably inferred that the underlying assumptions are grossly violated. In such cases, additional information needs to be obtained to better understand the Data Generating Process. Once this is obtained, an agreed mitigation strategy to adjust the data can be implemented. Note that no agreed mitigation strategy to adjust the data has been implemented.

2.4 Fitting discrete PDFs

Much of the process for fitting discrete probability distributions is similar to continuous distributions except that the goodness-of-fit statistic is assessed using the chi-square test[3]:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

Where O_i is the observed frequency for bin i and E_i is the estimated frequency. This is distributed with $k - p - 1$ degrees of freedom where p is the number outcomes estimated. Extreme outcomes can lead to unstable p-values; so in our application this process is simulated 2,000 times to compute a stable p-value.

The null hypothesis is that the observations are generated by the chosen distribution. This hypothesis is accepted if the chi-squared statistic is less than the prescribed threshold. This translates to a p-value higher than the p-value of the prescribed threshold.

2.5 Concluding remarks

For the reasons outlined in this section, Western Power has chosen the Anderson-Darling (**AD**) to explicitly reject poor fitting continuous statistical distribution from further consideration while the Chi-squared statistic is used to explicitly reject poor fitting candidate discrete statistical distributions. For the remaining candidate statistical distributions, the AIC is used to determine the best fitting statistical distribution for each performance measure.

[3] https://cran.r-project.org/doc/contrib/Ricci-distributions-en.pdf

3. Application

Key messages:

- The performance data is constructed according to the definitions as agreed for the AA5 period.
- SSBs are calculated according to the method described in Section 3.5.
- Last 5-year historical data is used to develop appropriate and stable performance measures.
- Mechanistically adjusting the performance measures risks seriously underestimating the true variance of the Data Generating Process. Setting SSBs using a percentile that is too low could lead to increased expenditure on network reliability to ensure compliance that is not valued by customers or provide a materially improved performance benefit for customers.
- The ERA Framework and Approach determined that the 97.5^{th} or 2.5^{th} percentile shall continue to be applied for the AA5 period to the performance measures that are subject to a statistical approach for the setting of the SSBs. That is, the measures of loss of supply event frequency, average outage duration, SAIDI, SAIFI and call centre performance.

  > *"The method for calculating the loss of supply event frequency, average outage duration, SAIDI, SAIFI and call centre performance benchmarks should continue to be based on the 97.5^{th} (or 2.5^{th}) percentile of actual performance over the previous period."*[4]

- The ERA Framework and Approach allows for Western Power to justify appropriate adjustments to the SSBs. This report does not include analysis for step change adjustments to the SSBs, as they are applied after the statistical methodology for the setting of the SSBs is applied.

3.1 Introduction

The method used by Western Power and the ERA to determine the SSBs for AA5 can broadly be described as follows:

1. Establish the data series relating to historic service standard performance (typically 5 years of 12 month rolling average data).
2. Determine the statistical distribution of best fit.
3. Sample the distribution to obtain the 97.5^{th} percentiles (or 2.5^{th} percentile where lower value reflects poorer performance), resulting in the SSB accordingly.
4. Where no statistical distribution was found to fit the data appropriately, the historic data was sampled directly.

The purpose of fitting the statistical distribution as opposed to simply sampling the history is to simulate the result should the dataset be larger, and to better model the probabilities in the tail end of the distribution.

3.2 Time period for fitting

As described above in section 2.2, this process assumes the history is relatively random and trend free, and that the future will be similar to the past. While true for some metrics, this does not always apply.

4 *ERA Framework and Approach for Western Power's fifth access arrangement review – Final Decision*, 9 August 2021, p. 33

In the past, the flaw in this assumption was addressed by using a shorter time series of data than available to fit the distribution, and some use of manual adjustments. Reducing the time period is effective at ensuring the mean of the distribution is more reflective (impacting the SST), however risks underestimating the variance or spread (impacting the SSB).

3.3 Selecting percentiles

Assuming stable performance, the sampling of the 97.5 th percentile should indicate a 2.5% probability of exceedance per metric. If the 12 AA5 SSBs (that use the statistical methodology) were fully independent, this would result in a 26.2% chance of exceeding at least one per year; effectively necessitating performance improvement to ensure compliance.[5] While the metrics are not fully independent, the impact is still valid.

In AA5, Western Power is proposing network investment to maintain service performance. The proposed network investment aligns closely with customer satisfaction analysis, indicating that customers are satisfied with the current level of performance.

3.4 Distribution selection

In this section, we give details about the SSB for each of 12 SSB measures using the method in section 2. A total of 16 different distributions were chosen (12 continuous and 4 discrete) based on their mathematical simplicity. This is consistent with the AA4 approach. The statistical distributions are listed in Table 1 below:

Table 1: Candidate statistical distributions

Continuous Distributions		Discrete Distributions
Exponential	3-Parameter Log logistic	Poisson
Gamma	Lognormal	Negative Binomial
3-Parameter Gamma	Normal	Binomial
Generalised Extreme Value	3-Parameter Lognormal	Geometric
Logistic	Weibull	
Log Logistic	3-Parameter Weibull	

The final statistical distribution for each service standard measure is selected based on the following **criteria**:

- Elimination of poorly fitting distributions. Each candidate distribution is tested to determine whether it should be rejected on the basis of the goodness-of-fit test, i.e., the Anderson-Darling (**AD**) test for continuous distribution and Chi-square test for integer distribution with p-value greater than 0.05. Note that the AD goodness-of-fit test can only be used for continuous distributions such as Normal and Weibull distributions, while the Chi-square test can be used for discrete distributions like Poisson and negative binomial distributions.

[5] This is calculated using the formula $P(A_1 \cup A_2 \ldots \cup A_n) = 1 - \sum_{i=1}^{n} P(A_i) - \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} P(A_i \cap A_i) + \sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1}\sum_{k=j+1}^{n} P(A_i \cap A_j \cap A_k) - \cdots$,where it is assumed to be no intersection among events, that is, $\sum_{1 \le i_1 < i_2 < \cdots < i_k \le n} P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = 0$

- Quality of fit. The distributions that have not been rejected by the relevant goodness-of-fit test are ranked according to the Akaike Information criterion (**AIC**). That is, the best fit is determined as having the lowest AIC.
- Stability of the resulting benchmark. Given that reliability data is continuously evolving, there is the risk that small changes in the underlying data would result in the AIC indicating a different distribution. The result can be a radical change in SSB benchmark from one year to the next, which would be contrary to the principles of benchmark setting.

The AIC is a measure of the relative quality of statistical distributions for the data. Given a collection of distributions for the same KPI data, AIC estimates the quality of each model, relative to each of the other distributions. There are two benefits by employing the AIC criterion: (1) discouraging overfit, and (2) overcoming the theoretical difficulty of comparing the AD or Chi-square statistics computed for several distributions fitted on the same KPI data. Hence, AIC provides a means for distribution ranking from a collection of fitted distributions (the smaller the AIC, the better the fit of the distribution).

All statistical distributions are implemented in R[6] with package **fitdistplus**.[7] The selection of distribution(s) is automatically selected by the first two of the above two. A comparison of fitted distribution parameters and associated quantiles are calculated and listed in Tables, and diagnostic plots for the final selected distributions are given per metric.

3.5 Summarised AA5 approach

The following proposed approach for setting the SSBs based on the statistical methodology has been implemented for each service standard measure based on the ERA Final Decision:

1. Establish the data series relating to historic service standard performance (5 years of 12 month rolling average data) based on AA5 Services Standards Definitions for the Western Power Network definitions [8];
2. Determine the statistical distribution of best fit based both an Anderson Darling (or Chi-square) test p-value above 0.05 and the lowest AIC
3. Sample the best fit distribution to obtain 97.5th percentiles (or 2.5th percentile where lower value reflects poorer performance) as the Services Standard Benchmark (SSB)
4. Calculate empirical 97.5th percentiles (or 2.5th percentile where lower value reflects poorer performance) as SSB in the case when there are no candidate distributions in Table 1 passing the goodness-of-fit test (p-value above 0.05) for a given historic service standard performance.

Note: Consistent with the ERA's Final Decision on the framework and approach, Service Standards Targets (SST) are set at the average annual level of performance achieved in the AA4 period, adjusted for anticipated changes in the service reliability and where individual penalty caps applied in the AA4period.

[6] A copy of the r code can be found at EDM # 43704994

[7] For more information about the fitdistplus package, see: https://cran.r-project.org/web/packages/fitdistrplus/index.html. Note that R packages on Cran are peer reviewed by experts.

[8] Section 4 of the proposed revisions to the Access Arrangement for the Western Power Network, EDM#56969357.

4. Distribution Network Performance Measures

For this proposal, the SSBs have been set based on the 97.5th percentile percentiles (or 2.5th percentile where lower value reflects poorer performance) of the distribution of best fit using historical performance data from July 2017 to September 2021. This analysis will be updated to use the full 5 years of AA4 actual historical performance data, July 2017 to June 2022, to set the final SSB targets for the AA5 period to apply from 2023/24 to 2026/27. This will be included into Western Power's response to the draft decision.

4.1 CBD SAIDI

All available distributions were fitted to data based on the AA5 proposed definition for the period July 2017 – September 2021. The performance statistics are shown in Table 2 below.

Table 2: Performance measures by candidate statistical distribution for CBD SAIDI

Distribution	Rank	Selected	AD p-value	AIC	p=0.975
Empirical			999	-999999	29.88
Weibull	1	Yes	0.242	366.05	35.23
Gamma	2	No	0.205	367.8	38.71
exp	3	No	0.063	373.62	48.36
Lognormal3	4	No	0.199	373.96	37.6
GEV	5	No	0.236	374.82	33.51
Lognormal	6	No	0.083	375.82	56.29
Normal	7	No	0.228	376.78	30.23
Weibull3	8	No	0.106	377.02	29.87
Loglogistic	9	No	0.099	377.64	71.61
Gamma3	10	No	0.102	378.88	30.23
Logistic	11	No	0.248	379.68	31.42
Loglogistic3	12	No	0.097	382.58	31.76

Table note: P indicates percentile

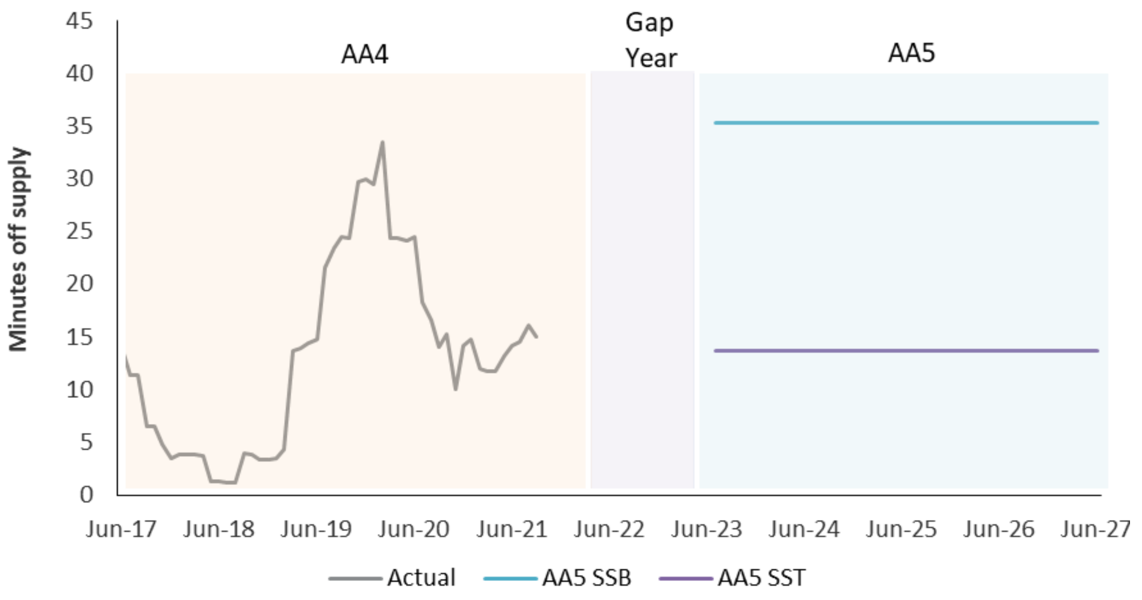
The process attempted to fit 12 statistical distributions to the proposed performance data based on the AA5 definition. Weibull distribution is the best fit distribution having both an Anderson Darling p-value above 0.05 and the lowest AIC. SSB is calculated as the 97.5th percentile from this distribution. SST is calculated as the average annual level of performance achieved in the AA4 period. The outcome is shown in Table 3 below:

Table 3: Proposed performance measures for CBD SAIDI

Scenario	Service Standard Target (SST)	Service Standard Benchmark (SSB)
AA5 proposed	13.7	35.2

Figure 3 below shows CBD SAIDI performance over the last 5 years in the AA4 period and proposed performance measures for AA5 period.

Figure 3: CBD SAIDI in the AA4 period and proposed performance measures in the AA5 period



4.2 Urban SAIDI

All available distributions were fitted to data based on the AA5 proposed definition for the period July 2017 - September 2021. The performance statistics are shown in Table 4 below.

Table 4: Performance measures by candidate statistical distribution for Urban SAIDI

Distribution	Rank	Selected	AD p-value	AIC	p=0.975
Empirical			999	-999999	136.16
Weibull3	1	Yes	0.97	373.9	138.9
Gamma3	2	No	0.962	375.64	140.36
Lognormal3	3	No	0.963	376.52	139.81
Lognormal	4	No	0.756	377	136.16
Gamma	5	No	0.711	377.5	135.88
Normal	6	No	0.601	378.72	135.46
Loglogistic3	7	No	0.933	379.22	144.68

Distribution	Rank	Selected	AD p-value	AIC	p=0.975
Loglogistic	8	No	0.764	379.56	137.5
Logistic	9	No	0.653	380.86	136.35
Weibull	10	No	0.218	387.86	134.96
GEV	11	No	0	538.01	138.65
exp	12	No	0	602.2	435.48

Table note: P indicates percentile

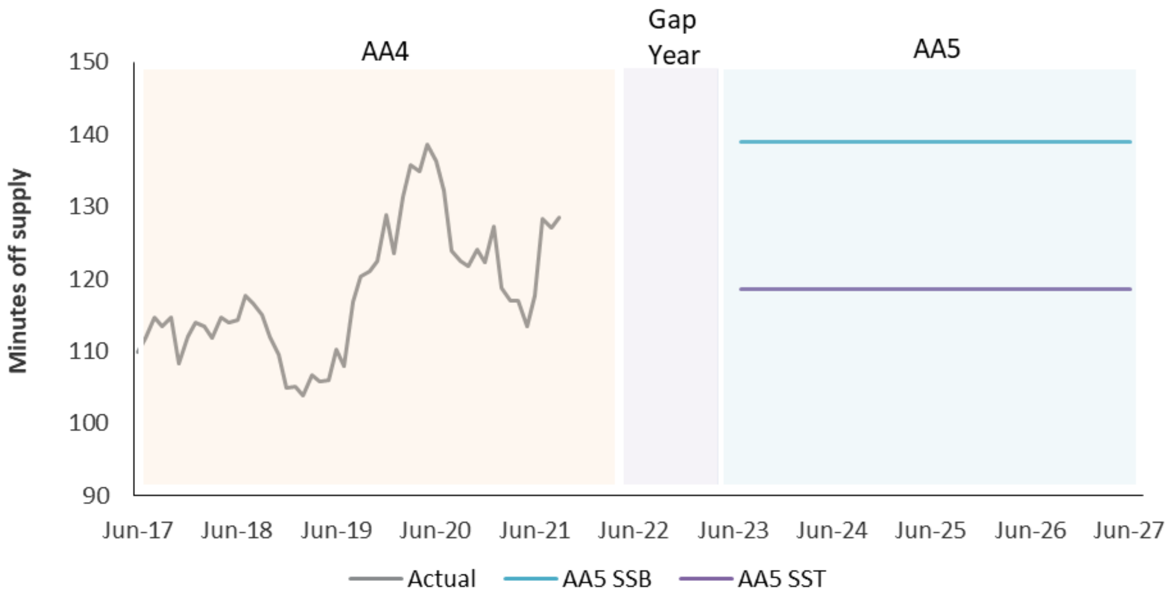
The process attempted to fit 12 statistical distributions to the proposed performance data based on the AA5 definition. Weibull3 distribution is the best fit distribution having both an Anderson Darling p-value above 0.05 and the lowest AIC. SSB is calculated as the 97.5 th percentile from this distribution. SST is calculated as the average annual level of performance achieved in the AA4 period with adjustment -0.7. The outcome is shown Table 5 below:

Table 5: Proposed performance measures for Urban SAIDI

Scenario	Service Standard Target (SST)	Service Standard Benchmark (SSB)
AA5 proposed	118.5	138.9

Figure 4 below shows Urban SAIDI performance over the last 5 years in the AA4 period and proposed performance measures for AA5 period.

Figure 4: Urban SAIDI in the AA4 period and proposed performance measures in the AA5 period



4.3 Rural Short SAIDI

All available distributions were fitted to data based on the AA5 proposed definition for the period July 2017 - September 2021. The performance statistics are shown in Table 6 below.

Table 6: Performance measures by candidate statistical distribution for Rural Short SAIDI

Distribution	Rank	Selected	AD p-value	AIC	p=0.975
Empirical			999	-999999	239.38
Weibull3	1	Yes	0.923	431.08	236.87
Gamma3	2	No	0.949	431.28	238.99
Lognormal3	3	No	0.945	432.57	243.78
Loglogistic3	4	No	0.912	434.52	260.56
Loglogistic	5	No	0.266	444.38	227.54
Lognormal	6	No	0.197	444.66	227.88
Gamma	7	No	0.172	446.18	227.63
Logistic	8	No	0.197	447.7	225.83
Normal	9	No	0.124	449.53	227.25
Weibull	10	No	0.034	465.2	229.93
GEV	11	No	0	614.22	239.56
exp	12	No	0	653.23	711.37

Table note: P indicates percentile

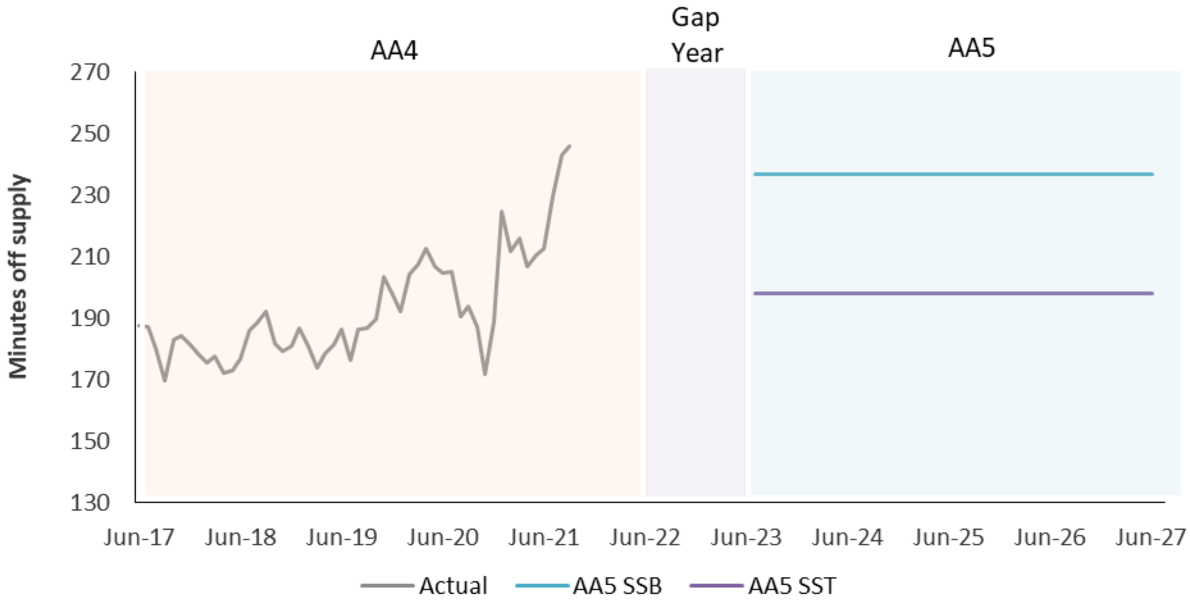
The process attempted to fit 12 statistical distributions to the proposed performance data based on the AA5 definition. Weibull3 distribution is the best fit distribution having both an Anderson Darling p-value above 0.05 and the lowest AIC. SSB is calculated as the 97.5 th percentile from this distribution. SST is calculated as the average annual level of performance achieved in the AA4 period with adjustment -0.6. The outcome is shown in Table 7 below:

Table 7: Proposed performance measures for Rural Short SAIDI

Scenario	Service Standard Target (SST)	Service Standard Benchmark (SSB)
AA5 proposed	197.9	236.9

Figure 5 below shows Rural Short SAIDI performance over the last 5 years in the AA4 period and proposed performance measures for AA5 period.

Figure 5: Rural Short SAIDI in the AA4 period and proposed performance measures in the AA5 period



4.4 Rural Long SAIDI

All available distributions were fitted to data based on the AA5 proposed definition for the period July 2017 - September 2021. The performance statistics are shown in Table 8 below.

Table 8: Performance measures by candidate statistical distribution for Rural Long SAIDI

Distribution	Rank	Selected	AD p-value	AIC	p=0.975
Empirical			999	-999999	805.88
Lognormal	1	Yes	0.829	570.3	812.5
Gamma	2	No	0.807	570.39	809.48
Normal	3	No	0.737	570.86	805.9
Weibull3	4	No	0.806	571.65	811.5
Loglogistic	5	No	0.827	571.89	821.07
Gamma3	6	No	0.841	572.25	814.41
Lognormal3	7	No	0.84	572.28	814.09
Logistic	8	No	0.751	572.49	812.53
Loglogistic3	9	No	0.858	573.75	828.95
Weibull	10	No	0.32	577.01	799.3
GEV	11	No	0	690.1	814.72
exp	12	No	0	786.67	2566.39

Table note: P indicates percentile

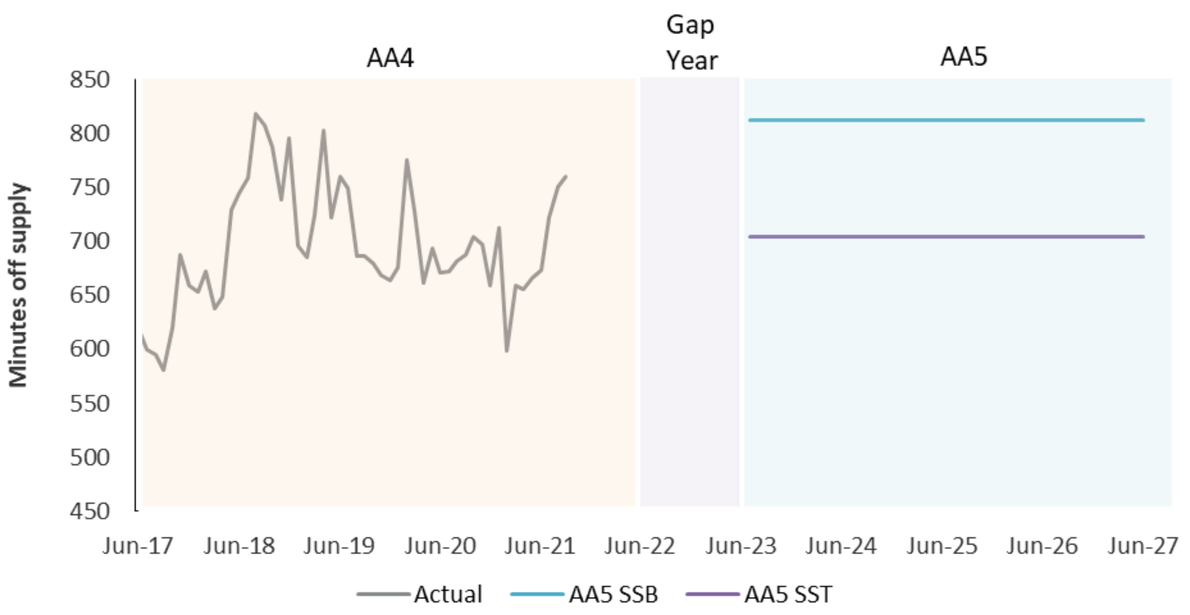
The process attempted to fit 12 statistical distributions to the proposed performance data based on the AA5 definition. Lognormal distribution is the best fit distribution having both an Anderson Darling p-value above 0.05 and the lowest AIC. SSB is calculated as the 97.5 th percentile from this distribution. SST is calculated as the average annual level of performance achieved in the AA4 period. The outcome is shown in Table 9 below:

Table 9: Proposed performance measures for Rural Long SAIDI

Scenario	Service Standard Target (SST)	Service Standard Benchmark (SSB)
AA5 proposed	704.3	812.5

Figure 6 below shows Rural Long SAIDI performance over the last 5 years in the AA4 period and proposed performance measures for AA5 period.

Figure 6: Rural Long SAIDI in the AA4 period and proposed performance measures in the AA5 period



4.5 CBD SAIFI

All available distributions were fitted to data based on the AA5 proposed definition for the period July 2017 - September 2021. The performance statistics are shown in Table 10 below.

Table 10: Performance measures by candidate statistical distribution for CBD SAIFI

Distribution	Rank	Selected	AD p-value	AIC	p=0.975
Empirical			999	-999999	0.31
Weibull3	1	Yes	0.065	-98.85	0.44
Weibull	2	No	0.111	-98.54	0.39
Gamma	3	No	0.119	-97.79	0.43
Gamma3	4	No	0.063	-97.64	0.44

Distribution	Rank	Selected	AD p-value	AIC	p=0.975
Lognormal	5	No	0.108	-95.48	0.53
Loglogistic	6	No	0.125	-90.43	0.66
GEV	7	No	0.087	-90.4	0.42
Normal	8	No	0.063	-89.19	0.36
Logistic	9	No	0.072	-83.37	0.38
Loglogistic3	10	No	0.058	-82.21	0.39
exp	11	No	0.015	-81.94	0.61
Lognormal3	12	No	0	-20.6	1331.01

Table note: P indicates percentile

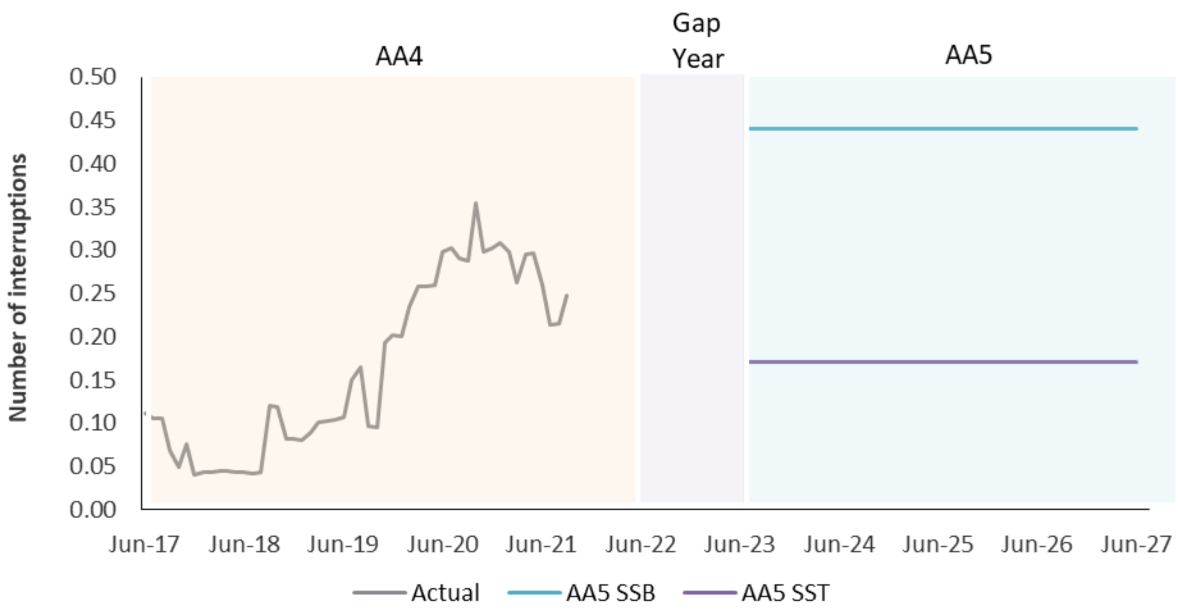
The process attempted to fit 12 statistical distributions to the proposed performance data based on the AA5 definition. Weibull3 distribution is the best fit distribution having both an Anderson Darling p-value above 0.05 and the lowest AIC. SSB is calculated as the 97.5 th percentile from this distribution. SST is calculated as the average annual level of performance achieved in the AA4 period with adjustment -0.02. The outcome is shown in Table 11 below:

Table 11: Proposed performance measures for CBD SAIFI

Scenario	Service Standard Target (SST)	Service Standard Benchmark (SSB)
AA5 proposed	0.17	0.44

Figure 7 below shows CBD SAIFI performance over the last 5 years in the AA4 period and proposed performance measures for AA5 period.

Figure 7: CBD SAIFI in the AA4 period and proposed performance measures in the AA5 period



4.6 Urban SAIFI

All available distributions were fitted to data based on the AA5 proposed definition for the period July 2017 - September 2021. The performance statistics are shown in Table 12 below.

Table 12: Performance measures by candidate statistical distribution for Urban SAIFI

Distribution	Rank	Selected	AD p-value	AIC	p=0.975
Empirical			999	-999999	1.33
Weibull	1	Yes	0.464	-124.56	1.33
Weibull3	2	No	0.574	-123.05	1.32
GEV	3	No	0.356	-121.01	1.33
Logistic	4	No	0.164	-115.33	1.38
Normal	5	No	0.099	-114.89	1.36
Loglogistic	6	No	0.121	-112.82	1.39
Gamma	7	No	0.076	-112.8	1.37
Lognormal3	8	No	0.094	-112.5	1.37
Gamma3	9	No	0.089	-112.02	1.37
Lognormal	10	No	0.067	-111.67	1.38
Loglogistic3	11	No	0	20.13	383.27
exp	12	No	0	126.07	4.47

Table note: P indicates percentile

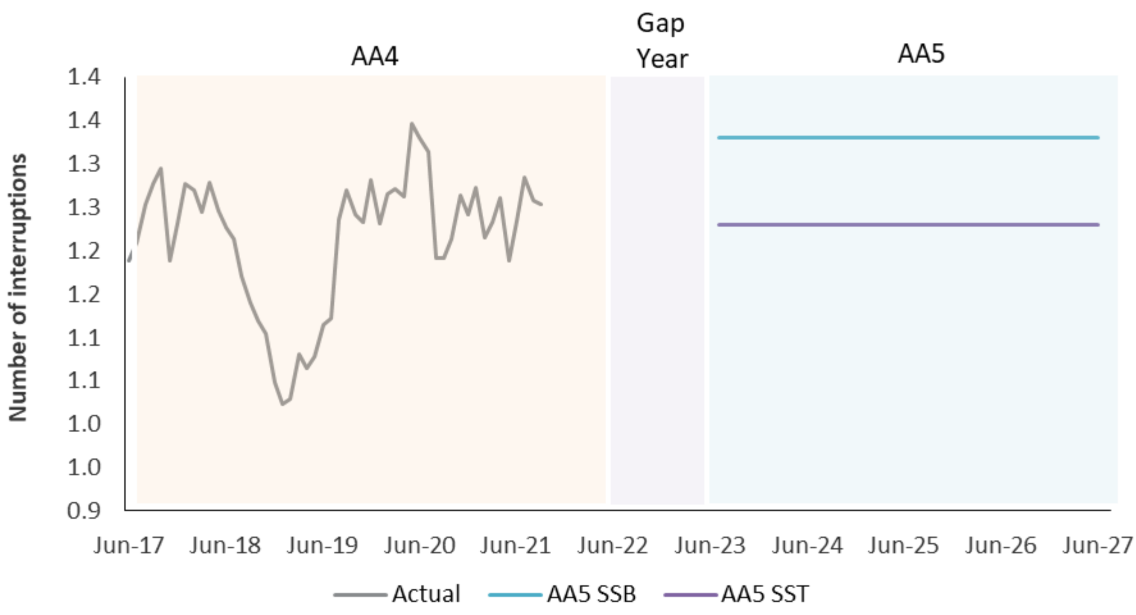
The process attempted to fit 12 statistical distributions to the proposed performance data based on the AA5 definition. Weibull distribution is the best fit distribution having both an Anderson Darling p-value above 0.05 and the lowest AIC. SSB is calculated as the 97.5th percentile from this distribution. SST is calculated as the average annual level of performance achieved in the AA4 period. The outcome is shown in Table 13 below:

Table 13: Proposed performance measures for Urban SAIFI

Scenario	Service Standard Target (SST)	Service Standard Benchmark (SSB)
AA5 proposed	1.23	1.33

Figure 8 below shows Urban SAIFI performance over the last 5 years in the AA4 period and proposed performance measures for AA5 period.

Figure 8: Urban SAIFI in the AA4 period and proposed performance measures in the AA5 period



4.7 Rural Short SAIFI

All available distributions were fitted to data based on the AA5 proposed definition for the period July 2017 - September 2021. The performance statistics are shown in Table 14 below.

Table 14: Performance measures by candidate statistical distribution for Rural Short SAIFI

Distribution	Rank	Selected	AD p-value	AIC	p=0.975
Empirical			999	-999999	2.24
Weibull3	1	Yes	0.76	-70.1	2.28
Gamma3	2	No	0.732	-68.68	2.31
Lognormal	3	No	0.428	-66.95	2.25
Gamma	4	No	0.412	-66.61	2.25
Lognormal3	5	No	0.479	-65.9	2.26
Normal	6	No	0.376	-65.79	2.24
Loglogistic	7	No	0.392	-62.92	2.28
Logistic	8	No	0.356	-62.02	2.26
Loglogistic3	9	No	0.428	-61.7	2.29
Weibull	10	No	0.221	-57.92	2.23
GEV	11	No	0	10.97	2.27
exp	12	No	0	178.05	7.38

Table note: P indicates percentile

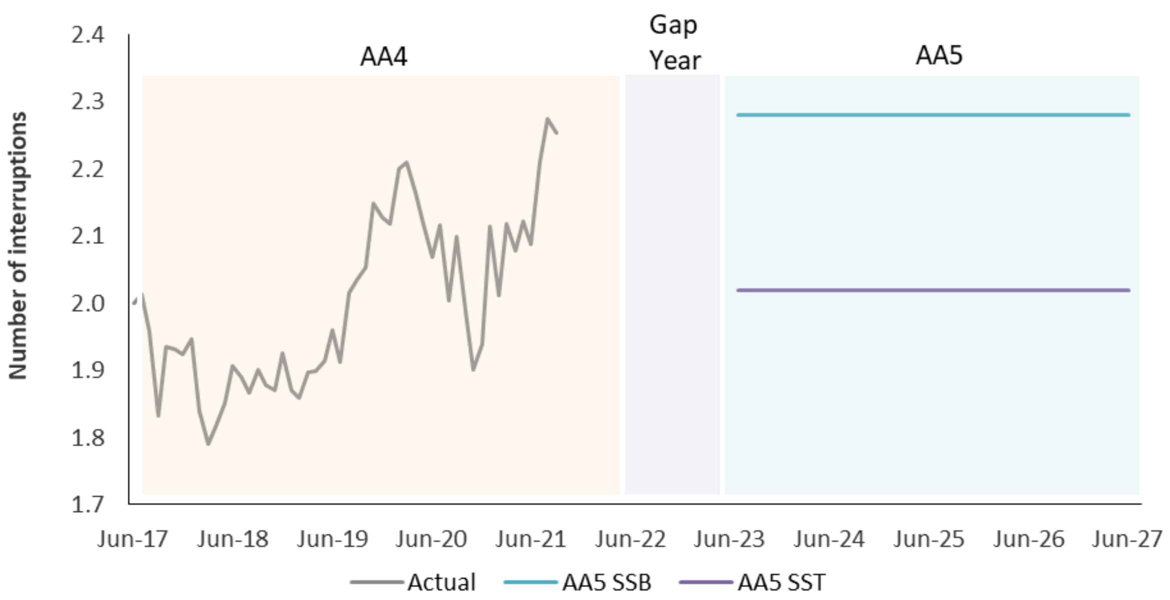
The process attempted to fit 12 statistical distributions to the proposed performance data based on the AA5 definition. Weibull3 distribution is the best fit distribution having both an Anderson Darling p-value above 0.05 and the lowest AIC. SSB is calculated as the 97.5 th percentile from this distribution. SST is calculated as the average annual level of performance achieved in the AA4 period. The outcome is shown in Table 15 below:

Table 15: Proposed performance measures for Rural Short SAIFI

Scenario	Service Standard Target (SST)	Service Standard Benchmark (SSB)
AA5 proposed	2.02	2.28

Figure 9 below shows Rural Short SAIFI performance over the last 5 years in the AA4 period and proposed performance measures for AA5 period.

Figure 9: Rural Short SAIFI in the AA4 period and proposed performance measures for Rural Short SAIFI



4.8 Rural Long SAIFI

All available distributions were fitted to data based on the AA5 proposed definition for the period July 2017 - September 2021. The performance statistics are shown in Table 16 below.

Table 16: Performance measures by candidate statistical distribution for Rural Long SAIFI

Distribution	Rank	Selected	AD p-value	AIC	p=0.975
Empirical			999	-999999	4.77
Loglogistic	1	Yes	0.903	-6.58	4.71
Logistic	2	No	0.893	-6.2	4.69
Lognormal	3	No	0.734	-5.65	4.7
Gamma	4	No	0.723	-5.46	4.69

Distribution	Rank	Selected	AD p-value	AIC	p=0.975
Normal	5	No	0.689	-4.92	4.68
Loglogistic3	6	No	0.896	-4.67	4.73
Lognormal3	7	No	0.73	-3.91	4.72
Gamma3	8	No	0.725	-3.88	4.71
Weibull3	9	No	0.613	-3.69	4.71
Weibull	10	No	0.153	6.31	4.68
GEV	11	No	0	103.89	4.82
exp	12	No	0	256.44	15.67

Table note: P indicates percentile

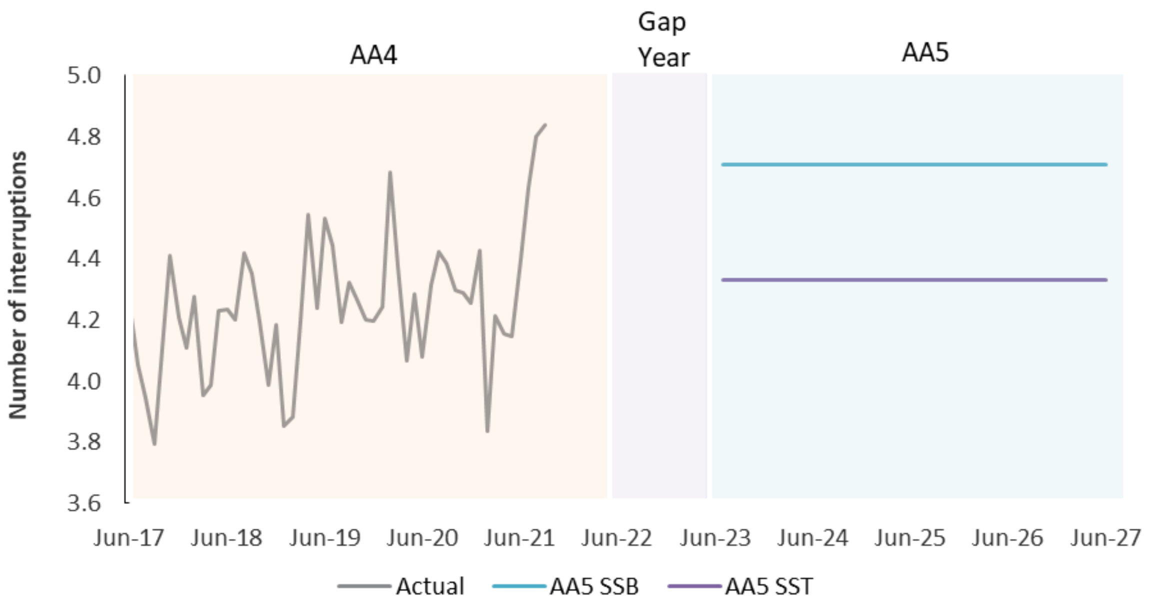
The process attempted to fit 12 statistical distributions to the proposed performance data based on the AA5 definition. Loglogistic distribution is the best fit distribution having both an Anderson Darling p-value above 0.05 and the lowest AIC. SSB is calculated as the 97.5 th percentile from this distribution. SST is calculated as the average annual level of performance achieved in the AA4 period. The outcome is shown in Table 17 below:

Table 17: Proposed performance measures for Rural Long SAIFI

Scenario	Service Standard Target (SST)	Service Standard Benchmark (SSB)
AA5 proposed	4.33	4.71

Figure 10 below shows Rural Long SAIFI performance over the last 5 years in the AA4 period and proposed performance measures for AA5 period.

Figure 10: Rural Long SAIFI in the AA4 period and proposed performance measures in the AA5 period



4.9 Call Centre Performance

All available distributions were fitted to data based on the AA5 proposed definition for the period July 2016 - September 2021. The performance statistics are shown in Table 18 below.

Table 18: Performance measures by candidate statistical distribution for Call Centre Performance

Distribution	Rank	Selected	AD p-value	AIC	p=0.025
Empirical			999	-999999	90.73
GEV	1	Yes	0.753	134.74	90.69
Gamma3	2	No	0.66	135.2	90.66
Lognormal	3	No	0.188	140.44	90.28
Gamma	4	No	0.184	140.57	90.27
Loglogistic	5	No	0.275	140.61	90.1
Normal	6	No	0.178	140.84	90.26
Logistic	7	No	0.264	140.92	90.08
Weibull	8	No	0.036	157.86	88.88
Weibull3	9	No	0.034	160.53	88.81
Loglogistic3	10	No	0	506.54	90.39
exp	11	No	0	576.3	2.33
Lognormal3	12	No	0	0	90.95

Table note: P indicates percentile

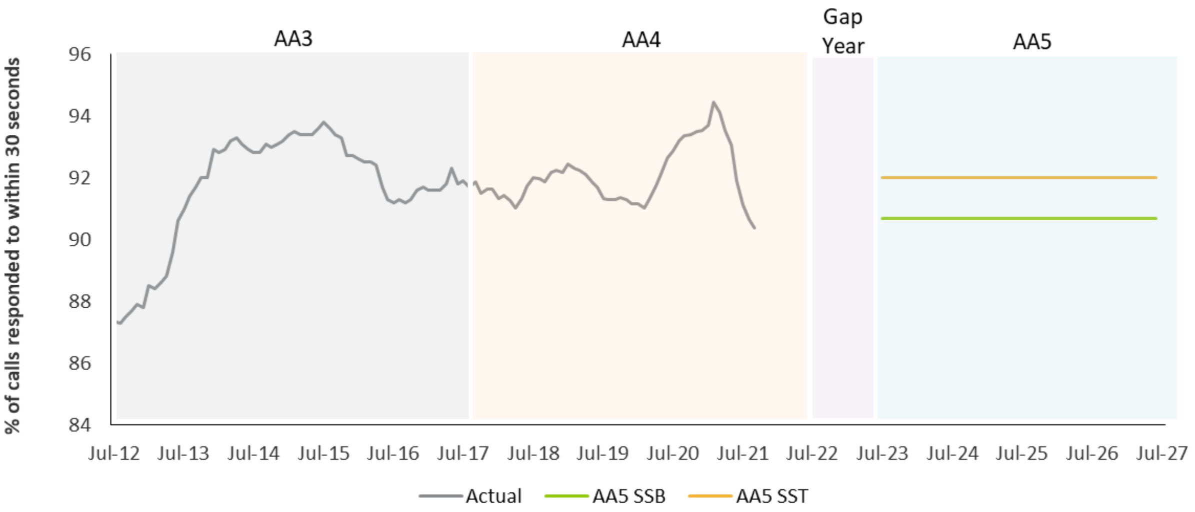
The process attempted to fit 12 statistical distributions to the proposed performance data based on the AA5 definition. GEV distribution is the best fit distribution having both an Anderson Darling p-value above 0.05 and the lowest AIC. SSB is calculated as the 2.5th percentile from this distribution. SST is calculated as the average annual level of performance achieved in the AA4 period. The outcome is shown in Table 19 below:

Table 19: Proposed performance measures for Call Centre Performance

Scenario	Service Standard Target (SST)	Service Standard Benchmark (SSB)
AA5 proposed	92.0%	90.7%

Figure 11 below shows Call Centre Performance over the last 5 years in the AA4 period and proposed performance measures for AA5 period

Figure 11: Call Centre Performance in the AA4 period and proposed performance measures in the AA5 period



Note the proposed AA5 SSB and SST for Call Centre in Figure 11 is based on the AA5 proposed definition for the period July 2016 - September 2021. This doesn't include the recent findings from the Customer and community engagement program (CEP). Further information on CEP findings and step-change to the Call Centre SSB & SST are provided in section 6.5, section 6.6 and section 6.7 of the Access Arrangement Information document.

5. Transmission Network Performance Measures

For this proposal, the SSBs have been set based on the 97.5th percentile percentiles (or 2.5th percentile where lower value reflects poorer performance) of the distribution of best fit using historical performance data from July 2017 to September 2021. This analysis will be updated to use the full 5 years of AA4 actual historical performance data, July 2017 to June 2022, to set the final SSB targets for the AA5 period to apply from 2023/24 to 2026/27. This will be included into Western Power's response to the draft decision.

5.1 Average Outage Duration

All available distributions were fitted to data based on the AA5 proposed definition for the period July 2016 - September 2021. The performance statistics are shown in Table 20 below.

Table 20: Performance measures by candidate statistical distribution for Average Outage Duration

Distribution	Rank	Selected	AD p-value	AIC	p=0.975
Empirical			999	-999999	1604.01
Gamma	1	Yes	0.313	766.29	1745.8
Weibull	2	No	0.265	767.22	1652.39
Lognormal3	3	No	0.299	767.92	1761.12
Gamma3	4	No	0.301	767.96	1710.46
GEV	5	No	0.285	768.02	1767.47
Loglogistic	6	No	0.26	768.62	2210.48
Weibull3	7	No	0.279	768.94	1665.12
Loglogistic3	8	No	0.267	769.58	1986.26
Lognormal	9	No	0.273	770.97	2081.5
Normal	10	No	0.134	775.19	1547.01
exp	11	No	0.001	795.85	2803.09
Logistic	12	No	0	843.81	1884.15

Table note: P indicates percentile

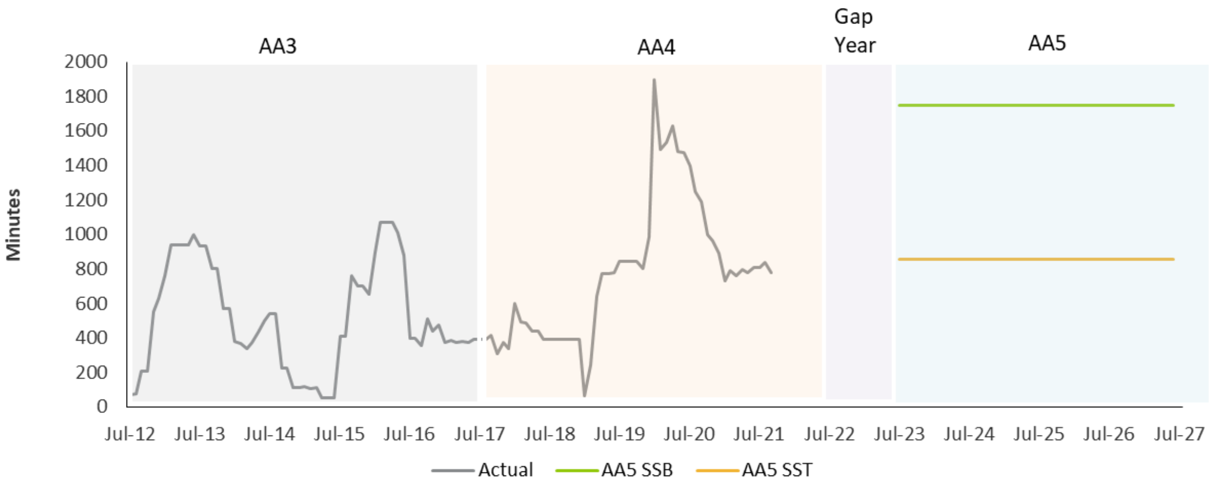
The process attempted to fit 12 statistical distributions to the proposed performance data based on the AA5 definition. Gamma distribution is the best fit distribution having both an Anderson Darling p-value above 0.05 and the lowest AIC. SSB is calculated as the 97.5 th percentile from this distribution and SST is calculated as the average annual level of performance achieved in the AA4 period. The outcome is shown in Table 21 below:

Table 21: Proposed performance measures for Average Outage Duration

Scenario	Service Standard Target (SST)	Service Standard Benchmark (SSB)
AA5 proposed	852	1746

Figure 12 below shows Average Outage Duration (**AOD**) performance over the last 10 years in the AA4 period and proposed performance measures for AA5 period

Figure 12: Average Outage Duration in the AA4 period and proposed performance measures in the AA5 period



5.2 Loss of Supply Event Frequency (>0.1 System Minutes Interrupted to ≤1.0 System Minutes Interrupted)

All available discrete distributions were fitted to the data based on the AA5 proposed definition for the period July 2016 - September 2021. The performance statistics are shown in Table 22 below:

Table 22: Performance measures by candidate statistical distribution for Loss of Supply Event Frequency (>0.1 System Minutes Interrupted to ≤1.0 System Minutes Interrupted)

Distribution	Rank	Selected	CS p-value	AIC	p=0.975
Empirical			999	-999999	4
Poisson	1	No	0.014	178.63	4
Negative binomial	2	No	0.02	178.82	5
Binomial	3	No	0.014	178.84	4
Geometric	4	No	0.001	182.94	7

Table note: P indicates percentile

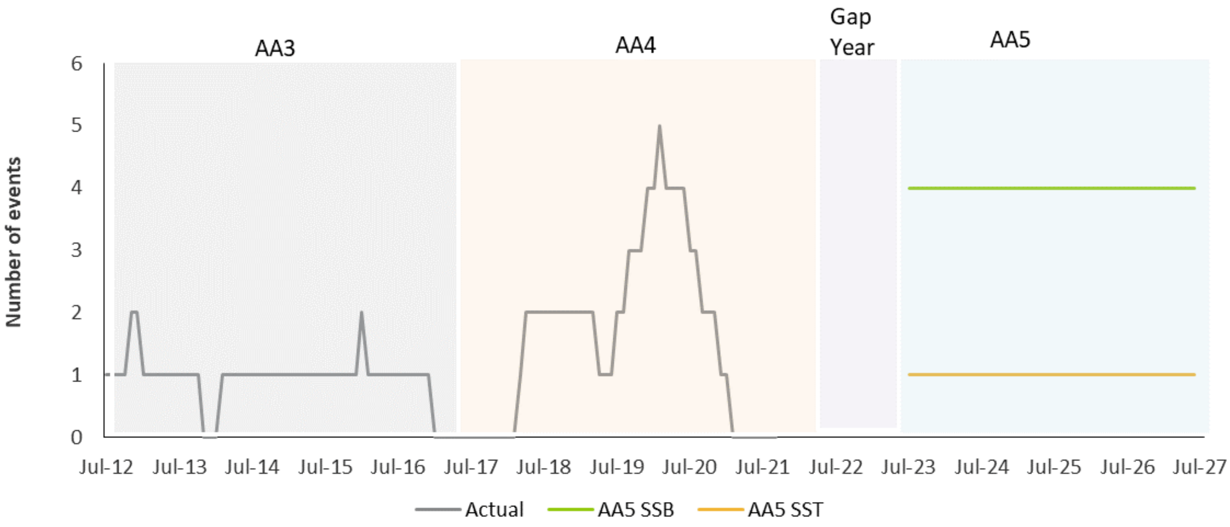
The process attempted to fit 4 discrete distributions to the proposed performance data based on the AA5 definition. Because there are no candidate distributions passing the goodness-of-fit test for 0.1 < Loss of Supply Event Frequency <= 1 (0.1<**LOSEF<=1**), the SSB is calculated by the 97.5th percentile. SST is calculated as the average annual level of performance achieved in the AA4 period. The outcome is shown in Table 23 below:

Table 23: Proposed performance measures for Loss of Supply Event Frequency (>0.1 System Minutes Interrupted to ≤1.0 System Minutes Interrupted)

Scenario	Service Standard Target (SST)	Service Standard Benchmark (SSB)
AA5 proposed	1	4

Figure 13 below shows Loss of Supply Event Frequency > 0.1 and ≤1.0 system minutes interrupted (**0.1<LOSEF≤1.0**) performance over the last 10 years in the AA4 period and proposed performance measures for AA5 period.

Figure 13: Loss of Supply Event Frequency (>0.1 System Minutes Interrupted to ≤1.0 System Minutes Interrupted) in the AA3-AA4 period and proposed performance measures in the AA5 period



5.3 Loss of Supply Event Frequency (>1 System Minute)

All available discrete distributions were fitted to the data based on the AA5 proposed definition for the period July 2016 - September 2021. The performance statistics are shown in Table 24 below:

Table 24: Performance measures by candidate statistical distribution for Loss of Supply Event Frequency (more than 1.0 System Minutes Interrupted)

Distribution	Rank	Selected	CS p-value	AIC	p=0.975
Empirical			999	-999999	2
Binomial		No	0.013	105.85	3
Poisson		No	0.013	106.02	3
Negative binomial		No	0.012	108.02	3
Geometric		No	0	121.07	4

Table note: P indicates percentile

The process attempted to fit 4 discrete distributions to the proposed performance data based on the AA5 definition. Because there are no candidate distributions passing the goodness-of-fit test for Loss of Supply

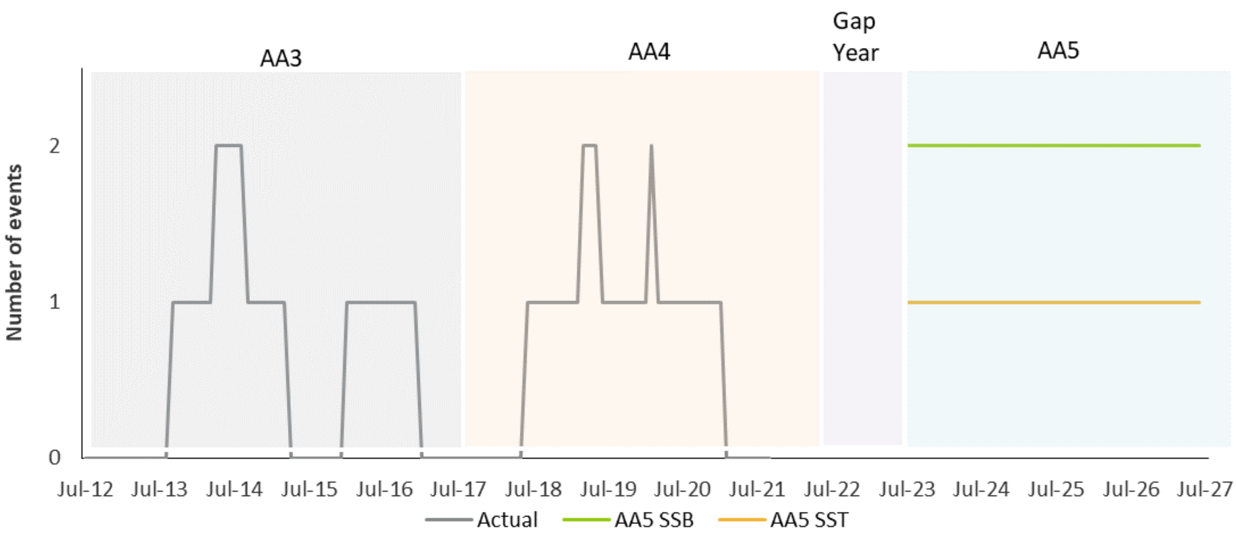
Event Frequency > 1 (**LOSEF>1**), the SSB is calculated by the 97.5th percentile. SST is calculated as the average annual level of performance achieved in the AA4 period. The outcome is shown in Table 25 below:

Table 25: Proposed performance measures for Loss of Supply Event Frequency (more than 1.0 System Minutes Interrupted)

Scenario	Service Standard Target (SST)	Service Standard Benchmark (SSB)
AA5 proposed	1	2

Figure 14 below shows Loss of Supply Event Frequency > 1 (**LOSEF>1**) performance over the last 10 years in the AA4 period and proposed performance measures for AA5 period

Figure 14: Loss of Supply Event Frequency (more than 1 minute) in the AA3-AA4 period and proposed performance measures in the AA5 period



6. References

Castro, R 2013: Lectures 2 and 3 – Goodness-of-Fit (**GoF**) Tests; URL: https://pdfs.semanticscholar.org/1b85/f2b451b8bbcbc9e5cfbfe317b0fd2464b2e9.pdf

Deligenett-Muller, M. L. and C. Dutang, 2015: fitdistplus: An R Package for Fitting Distributions. *Journal of Statistical Software*. Vol 64, Issue 4, 1-34.

Western Power Report, 2015: Reliability KPIs-Setting the Service Standard Benchmark (**SSB**) and Service Standard Target (**SST**).

Western Power Report, 2017: Reliability KPIs-Setting the Service Standard Benchmark (**SSB**) and Service Standard Target (**SST**).