Comments on Chapter 8 of the Commerce Commission Discussion Paper

Reset of Default Price-Quality Path for Electricity Distribution Businesses (19 June 2009)

by

Peter Thomson
Statistics Research Associates Ltd
Wellington, New Zealand

17 July 2009
Revised 30 July 2009

Summary of recommendations

This commentary is focussed exclusively on Chapter 8 (Quality Standards) of the New Zealand Commerce Commission discussion paper ComCom (2009). It is informed by the results to date of ongoing research undertaken by Statistics Research Associates on the quality of supply by New Zealand electricity distribution networks. In addition to other more general comments, it makes the following specific recommendations.

Recommendation 1 *With regard to the normalisation of daily SAIDI and SAIFI values, it is recommended that*

(a) estimates of the threshold T_{MED} be based on the non-normalised daily SAIDI and SAIFI values in the five year reference period;

(b) normal variation be assessed using truncated daily data (MEDs excluded) in keeping with IEEE standard 1366^{TM}-2003;

(c) the assumption that non-zero daily SAIDI or SAIFI values follow a log-normal distribution be replaced by the assumption that they follow a more flexible family of distributions such as a mixture of log-normal distributions;

(d) zero SAIDI values be accounted for in determining T_{MED}.

Recommendation 2 *With regard to dead-bands it is recommended that*

(a) dead-bands be constructed using the principles of statistical hypothesis testing with daily data from the reference period used to establish the hypothesis of compliance;

(b) the false-positive rate α, the probability a network is deemed to be in breach of its quality standard when it is in fact compliant, be set at a common value for all networks in keeping with the fairness principle of IEEE standard 1366^{TM}-2003.

Recommendation 3 *With regard to assessing compliance over time, it is recommended that*

(a) procedures such as the simple "two in a row" approach be adopted to enhance transparency while retaining flexiblity in setting dead-band upper limits and controlling false-positive rates;

(b) the false-positive rates of any competing approaches be explicitly evaluated and compared.

Further details are given in the following sections.

1 Preamble

On 19 June 2009, the New Zealand Commerce Commission (Commission) released a discussion paper (ComCom 2009) in which it sets out its preliminary views in relation to the reset of the default price-quality path that expires on 31 March 2010. The Commission has invited submissions on the discussion paper which should be received by the Commission before 5pm on Friday 17 July 2009.

Given the short time frame, these comments are brief and are focussed exclusively on Chapter 8 (Quality Standards) of ComCom (2009). They are informed by the results to date of ongoing research on the quality of supply by New Zealand electricity distribution networks. This research, undertaken by Statistics Research Associates (SRA) and commissioned by the New Zealand Electricity Networks Association (ENA), is concerned with establishing a robust statistical framework for measuring performance and setting performance targets for quality of supply. Much of the current focus of the research has been on the statistical properties of the daily SAIDI (System Average Interruption Duration Index) and SAIFI (System Average Interruption Frequency Index) reliability indices for New Zealand electricity distribution networks. The Commission was provided with a progress report on this research on 8 April 2009.

In ComCom (2009) the Commission adopts the important principle of "no material deterioration" for assessing service quality and the use of SAIDI and SAIFI reliability measures. The Commission also considers a quality assessment framework with the following key attributes.

Reference period The SAIDI and SAIFI indices over a five year historical period (*reference period*) from 1 April 2004 to 31 March 2009 are to be used to establish benchmark quality standards against which to assess "no material deterioration". In the subsequent five year *regulatory period*, the performances of SAIDI and SAIFI in each successive *assessment year* are to be compared with the corresponding performances in the reference period.

Quality standards The benchmark quality standards to be used are the averages of the SAIDI and SAIFI values over the reference period, where these averages are calculated from daily SAIDI and SAIFI values adjusted for extremes (*normalisation*).

Assessment At the end of any annual assessment year, the averages (or totals) of the normalised daily SAIDI and SAIFI values for that year (*quality measures*) are compared to the corresponding averages (*quality standards*) determined from the reference period. The electricity distribution network concerned would be assessed as compliant if there were no sustained deterioration in quality between the assessment year and the reference period, and non-compliant otherwise.

Normalisation To reduce the effects of rare extreme events on daily SAIDI and SAIFI averages, the daily SAIDI and SAIFI values concerned are adjusted (normalised) using a variant of the IEEE standard 1366^{TM}-2003 (IEEE 2004), subsequently recognized as an American National Standard (ANSI).

Dead-bands To account for sampling variation in the annual averages of (normalised) daily SAIDI and SAIFI values over any assessment year, network dependent dead-bands are to be introduced that account for this variability. If a network's *quality measure* falls above its dead-band, then this would constitute a breach of the quality standard.

Assessing compliance It is possible that a network will breach its quality standards when the underlying process generating its daily SAIDI and SAIFI values is compliant. To reduce the occurrence of technical breaches such as these, assessment is to be undertaken using *assessed values* where these are based on the performance of the quality measures (SAIDI and SAIFI) over the last three years. It is suggested that the assessed value of any quality measure be defined as the smaller of its value in the current assessment year, and the average of its values in the current and two previous years (three year average). Assessed values would then be compared to their corresponding quality standards.

In keeping with the objective of identifying sustained, rather than transitory, deterioration over time, a network would be considered in breach of compliance in any assessment year if its assessed values failed to meet their quality standards in the current year and at least one of the previous two years.

This framework will be discussed in the following sections.

2 Reference period

The Commission have adopted a process where quality standards for electricity distribution networks are normally reset every five years. In any five year regulatory period, a network's SAIDI and SAIFI reliability measures over each successive assessment year are assessed against the corresponding measures over the five year reference period preceding the regulatory period. The Commission may choose to extend a regulatory period, as has been done with the current period, but the roughly five year cycle would appear to be a continuing feature of the Commission's approach.

An alternative process would be annual resets where a network's annual SAIDI and SAIFI reliability measures are assessed against the corresponding measures in the immediately

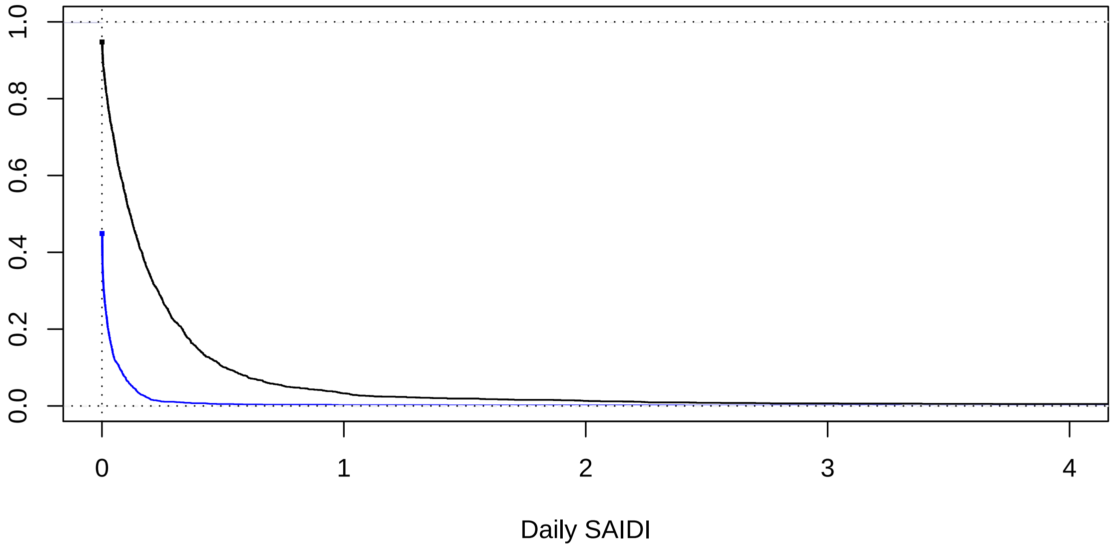

Figure 1: Empirical survivor functions of daily SAIDI values for Auckland (black) and Wellington (blue) over the period 1 April 2002 to 31 March 2007. One Wellington and 9 Auckland daily SAIDI values greater than 4 minutes are not shown.

preceding five years. Such a procedure is a rolling annual assessment where the reference period is no longer fixed, but defined with reference to the assessment year. This process would appear to be the regulatory cycle implied by the IEEE standard 1366^{TM}-2003 (IEEE 2004) and its supporting documentation.

3 Quality standards

The daily SAIDI and SAIFI reliability measures considered are non-negative and the smaller they are the better. Over any given period, the performance of such a daily measure is characterised by the proportion of days when no sustained outages occurred, and the distribution of its non-zero values.

Figure 1 shows examples of the distribution of daily SAIDI values for the Auckland and Wellington electricity distribution networks over the period 1 April 2002 to 31 March 2007. The distribution of sample data can be summarised in many equivalent ways (histogram, empirical distribution function, empirical survivor function etc). Here the empirical survivor function

$$\hat{S}(x) = \text{proportion of daily SAIDI values exceeding } x \qquad (x > 0)$$

has been plotted since it clearly illustrates the important role played by the zero SAIDI days and, in addition, the average daily SAIDI value over the 5 year period (including zero SAIDI days) is given by the area under $\hat{S}(x)$.

From Figure 1 it is readily seen that the two networks have markedly different distributions which reflect their very different scales, topologies and environments. The mean daily

SAIDI value for Wellington (0.028 minutes) is an order of magnitude smaller than that for Auckland (0.260 minutes), and the proportion of zero SAIDI days for Wellington (55.1%) is an order of magnitude higher than that for Auckland (5.2%). The plots show clear evidence of long (heavy) upper tails with 99% of Wellington's daily SAIDI values being less than 0.30 and 99% of Auckland's values being less than 2.24. Indeed, since the horizontal axis of the plot was clipped at four minutes for presentation purposes, one Wellington and nine Auckland daily SAIDI values greater than four minutes are not shown.

The ideal survivor function would be identically zero when all days in the period concerned are zero SAIDI days. In general, the closer $\hat{S}_t(x)$ is to zero the better. In practice this can be achieved by increasing the proportion of zero SAIDI days, or shrinking the distribution of non-zero SAIDI values to zero, or both.

In terms of measuring reliability performance, the proportion of zero SAIDI days is evidently an important factor, as is the shape (upper tail in particular) of the survivor function. Many summary measures of the shape are possible, with the area under the survivor function (the average daily SAIDI value over the period) being just one. Note, however, that there are many different survivor functions that have the same mean, but substantially different shapes (the only requirement is that the area under each survivor function is the same).

The survivor function $\hat{S}_+(x)$ of the non-zero SAIDI days is given by

$$\hat{S}_+(x) = \hat{S}(x)/(1 - \hat{p}_0)$$

where $\hat{p}_0$ is the proportion of zero SAIDI days. Since $\hat{S}_+(x)$ is just a rescaled version of $\hat{S}(x)$, the mean of the non-zero SAIDI days (the area under $\hat{S}_+(x)$) is just the overall mean divided by the proportion of non-sero SAIDI days or, equivalently,

$$\text{mean of all SAIDI values} = (1 - \hat{p}_0)(\text{mean of the non-zero SAIDI values}).$$

This is a key relationship. Instead of measuring reliability performance using just the mean of all SAIDI values (including zero SAIDI days), one could measure performance using the proportion of zero SAIDI days $\hat{p}_0$ and the mean of the non-zero SAIDI values (two separate measures rather than just one).

The Commission considers that averages of SAIDI and SAIFI values over the reference period are suitable measures to benchmark quality standards. Whatever the averages adopted, it is clearly important to take account of zero SAIDI days and be aware of the special role that they play.

4 Assessment

At the end of any annual assessment year, the total annual SAIDI or SAIFI value is to be calculated for that assessment year, and compared to the average of the total annual SAIDI or SAIFI values for each of the five years in the reference period. In all cases, a total annual SAIDI or SAIFI value is the sum of the (normalised) daily SAIDI or SAIFI values over all days in the year. The electricity distribution network concerned would be assessed as compliant if there were no sustained deterioration in quality between the assessment year and the reference period, and non-compliant otherwise.

Alternatively, this procedure can be recast as a comparison between the average daily SAIDI or SAIFI value over the assessment year and the same average over the reference period. If the average over the assessment year is significantly greater than the average over the reference period, then the network would be deemed non-compliant and compliant otherwise. The two comparisons (involving average annual totals on the one hand, and average daily values on the other) are equivalent and nothing is lost by using either. However the alternative interpretation focuses attention on the distribution of daily SAIDI and SAIFI values which is instructive for statistical modelling and inference.

Unless otherwise stated, the alternative interpretation of the assessment criterion framed in terms of averages of daily SAIDI or SAIFI values will be assumed in what follows.

In essence, the distribution of daily SAIDI or SAIFI values over the assessment year is being compared to the distribution of daily values over the reference period. If the distribution of daily SAIDI or SAIFI values for the assessment year is much the same as, or closer to zero than, the comparable distribution for the reference period, then this would constitute evidence of compliance. There are many statistical techniques (hypothesis testing in particular), methods and measures that could be used to make this comparison, with the comparison of means being just one of many possibilities.

Whatever assessment measure is adopted, it will always be assumed that the statistical principle of comparing like with like is always adhered to (comparing means of normalised data from the assessment year to means of normalised data from the reference period for example).

5 Normalisation

Like their overseas counterparts, the Commission recognises that daily SAIDI and SAIFI values have distributions that are typically heavy-tailed and, as a consequence, can result in rare extreme values that significantly distort the average of daily values over any assessment year. These large values are caused by rare events (often related to weather) that have occurred infrequently in the past, and will continue to occur infrequently in the future. These events are sufficiently rare that they fall outside the design specifications of individual electricity distribution networks and, as a consequence, give rise to daily SAIDI or SAIFI values that are unrepresentative of their normal day-to-day operations.

To reduce the impact of such rare extreme events on annual averages of daily SAIDI and SAIFI values, the Commission proposes adjusting or normalising daily reliability data using a variant of the IEEE standard 1366^{TM}-2003 (IEEE 2004). This variant is described in ComCom (2007) and based on recommendations made in PBA (2007).

5.1 IEEE Beta method

Two key principles underpin the so-called Beta Method described in IEEE (2004). These are as follows.

Classification Each daily SAIDI value is classified as either a *Major Event Day* (MED), during which a network's reliability is stressed beyond that normally expected, or a *normal day*. A daily SAIDI value falls in one or other category depending on whether it exceeds, or doesn't exceed, a suitably chosen threshold T_{MED}.

Fairness The Major Event Day (MED) identification threshold T_{MED} is chosen to be an *agreed common percentile* of each network's daily SAIDI distribution. Here T_{MED} is defined to be the 99.38% percentile of a network's daily SAIDI distribution. Although the values of T_{MED} will be network dependent, the same proportion of MEDs (0.62%) will be identified for each network, on average, over any given period of time. This underpins the *fairness criterion*.

At the end of each annual reporting period, a network is expected to fully report on all MEDs to the regulator, but only provide a suitable summary of the normal daily performance (a total annual SAIDI value with MEDs excluded).

The Beta Method allows both the networks and the regulator to study MEDs separately from normal day-to-day variation. The former are reviewed, possibly more rigorously, to determine their likely cause, whether the rate of MEDs is increasing, and what steps might be taken to reduce the incidence of MEDs in the future. On the other hand, the summary measures of normal day-to-day performance can be benchmarked against suitable quality standards, and to check for trends that might otherwise be obscured by the large MED values. The classification of daily SAIDI data into two distinct categories without modification is an important aspect of the Beta Method; no data is created or excluded. Instead the data in each classification is reported on using separate procedures.

The IEEE standard 1366^{TM}-2003, subsequently recognized as an American National Standard (ANSI), introduces a consistent means of defining a rare extreme event based on whether a daily SAIDI value exceeds a threshold T_{MED} or not. However, despite its attractions, relatively few regulators have adopted it. In the US, for example, a survey of 51 state regulators in 2006 found that only four used IEEE standard 1366^{TM}-2003 (see Eto and LaCommare 2008). Although the general principles that underpin the Beta Method appear to have been largely accepted, there has been considerable debate and discussion concerning the estimation of the threshold T_{MED}.

In practice, extreme quantiles such as T_{MED} are very difficult to estimate accurately. They require observations of sufficient number and quality together with a statistical model that fits the data well. The Beta Method assumes that non-zero daily SAIDI values follow a log-normal distribution which can then be used to estimate T_{MED}. This assumption has been criticised for the following reasons.

Failure to take proper account of zero SAIDI days Many research papers and IEEE Working Group reports (see Christie 2003a, 2003b, 2003c, Warren and Saint 2005, for example) have addressed this issue, but so far without leading to a change in IEEE standard 1366^{TM}-2003 which excludes zero SAIDI days in the calculation of T_{MED}. In essence, the Beta Method assumes that the number of zero SAIDI days has little impact on the calculation of T_{MED} and, more importantly, the annual totals of normal daily SAIDI values where MEDs are excluded.

Assumption that non-zero daily SAIDI values follow a log-normal distribution
A number of US utilities argued that their daily SAIDI data did not follow a log-normal distribution, with some even reporting bi-modal distributions (which cannot be log-normal). Christie (2003a) used graphical statistical methods (quantile-quantile or QQ plots) to show that, for the particular daily SAIDI data set considered, log-normality was a reasonable assumption. However others such as Billinton and Acharya (2006) contest the general applicability of this result to all utilities.

As shown in Christie (2003a) and as indicated earlier in Section 3, taking account of zero SAIDI days is not difficult conceptually, nor a major technical complication. Furthermore, if the log-normal fails to provide a good approximation to the distribution of non-zero daily SAIDI values, then there are other more general families of distributions that can.

5.2 Commission's variant of the Beta method

The Commission's proposed variant of the Beta Method is described in ComCom (2007) and based on recommendations made in PBA (2007). However the recommendations made in the latter are based on a flawed analysis and, in addition, deviate from the principles underpinning IEEE standard 1366^{TM}-2003. In particular PBA (2007)

(a) propose a Beta Method that differs from that given by IEEE standard 1366^{TM}-2003 since it estimates T_{MED} from past data with both MEDs and zero SAIDI days excluded;

(b) assess normal variation using a total annual SAIDI value with MEDs replaced by the threshold value T_{MED} (a process known as censoring in statistics) rather than excluding MEDs as stipulated in IEEE standard 1366^{TM}-2003;

(c) conclude that New Zealand networks have non-zero daily SAIDI values that follow a log-normal distribution, where the supporting analysis is deficient and apparently incorrect (the plots are inadequate and not QQ plots, standard statistical goodness-of-fit tests are not applied; confidence intervals and the sampling variability of estimates of T_{MED} are incorrect);

(d) fail to properly account for the significant numbers of zero SAIDI days that would appear to be a feature of New Zealand's small and geographically diverse networks.

With regard to point (a), the IEEE standard 1366^{TM}-2003 estimates T_{MED} for any given assessment year from the non-normalised daily SAIDI values over the immediately preceding five year reference period. However, as noted in Section 2, the Commission has adopted a five year regulatory cycle within which the five year reference period is fixed. To conform to this cycle it would be appropriate, simpler, and less costly, to base the calculation of T_{MED} on the non-normalised daily SAIDI or SAIFI values in the fixed five year reference period associated with each regulatory period.

With regard to point (b), the censoring of MEDs is at variance with the first of the two underpinning principles of IEEE standard 1366^{TM}-2003 which requires that daily SAIDI values be classified into two distinct categories (MEDs and normal variation) without modification. If MED values are replaced by large thresholds, many of the variability issues inherent in the original data will remain in the censored data and, in particular, may be difficult for networks to manage on a day-to-day basis. Such issues, in turn, are likely to impact normal operational practice (scheduling of preventive maintenance, for example) to the detriment of customers.

Censoring also places much greater importance on the estimation of the threshold T_{MED} than the IEEE Beta method where MEDs are omitted (referred to as truncation in statistics). If an estimate of T_{MED} is too high (as seems likely for New Zealand data if the log-normal assumption is adopted) then this biases the mean of the normal variation upwards and introduces unnecessary additional variability for the networks to manage.

With regard to points (c) and (d), the analysis of the New Zealand electricity reliability data undertaken by SRA to date shows that

- the assumption that non-zero daily SAIDI values follow a log-normal distribution does not hold for New Zealand network data, where formal statistical goodness-of-fit tests overwhelmingly reject this hypothesis;

- zero SAIDI days are a major issue for many New Zealand networks that is not dealt with properly in IEEE (2004), nor in PBA (2007).

These issues are illustrated in Figures 2 and 3.

If the non-zero daily SAIDI values follow a log-normal distribution, then the logarithms of the non-zero daily SAIDI values follow a normal distribution. Figure 2 shows the histograms of the logarithms of the non-zero daily SAIDI values for Auckland and Wellington over the period 1 April 2002 to 31 March 2007 with best fitting normal distributions (red) superimposed. The best fitting normal distribution is clearly inadequate in both cases, especially in the upper tail which is used to determine T_{MED}.

By contrast, Figure 2 shows that the best fitting mixture of two normal distributions (blue) yields a much more acceptable fit to both data sets and, in particular, the upper tails. It is noted that a mixture of two normal distributions provides a very flexible family of distributions that includes the normal distribution as a special case. It can account for bimodality as well as skewness and is widely used in practice to provide accurate, flexible approximations of data distributions. Furthermore, the model is readily fitted to data, and its component normal distributions often admit useful interpretations.

Adopting such a family for the logarithms of non-zero daily SAIDI values is equivalent to replacing the log-normal assumption by the more general assumption that non-zero SAIDI values follow a mixture of two log-normal distributions. If Figure 2 is typical of

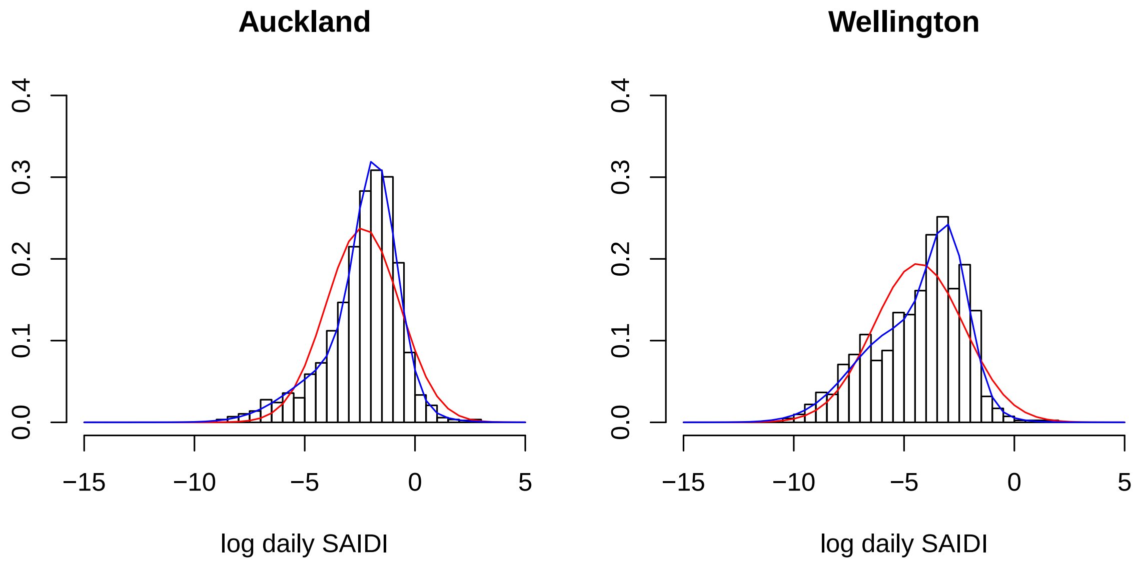

Figure 2: Histograms of the logarithms of the non-zero daily SAIDI values for Auckland and Wellington over the period 1 April 2002 to 31 March 2007. A best fitting mixture of two normal distributions (blue) and the best fitting normal distribution (red) are superimposed.

the New Zealand experience, then such an assumption is much more likely to hold over all New Zealand networks. This is a focus of the ongoing research undertaken for ENA by SRA. However it will always be true that a log-normal mixture distribution will fit better than the log-normal distribution since the latter is just a special case of the former.

The impact on T_{MED} of both the choice of distribution (log-normal or log-normal mixture) and the need to account for zero SAIDI days is clearly evident in Figure 3. Here the empirical survivor functions of daily SAIDI values for Auckland and Wellington over the period 1 April 2002 to 31 March 2007 are plotted, together with the survivor functions of the best fitting log-normal (including and excluding zero SAIDI days), and the best fitting log-normal mixture (adjusted for zero SAIDI days). The corresponding values of T_{MED} are also shown together with the empirical 99.38% percentile determined directly from the data (black). The log-normal mixture, adjusted for zero SAIDI days, produces the best fit to both data sets and produces values of T_{MED} that most closely approximate the empirical 99.38% percentile. There is little impact on the log-normal fits when the proportion of zero SAIDI days is small (5% for Auckland), but makes a dramatic difference when the proportion of zero SAIDI days is large (55% for Wellington). For both Auckland and Wellington, the best estimates of T_{MED} obtained from the log-normal mixture, or the empirical 99.38% percentile, are substantially less than those determined using the log-normal with zero SAIDI days excluded (approximately 50% less in the case of Auckland and 80% less in the case of Wellington).

Figure 3 highlights the need to adopt a suitable family of distributions for modelling non-zero SAIDI days, and the need to account for zero SAIDI days. Otherwise, the principle of using a common percentile of the daily SAIDI distribution to classify MEDs across networks is likely to be violated, possibly strongly. This brings into question the fairness

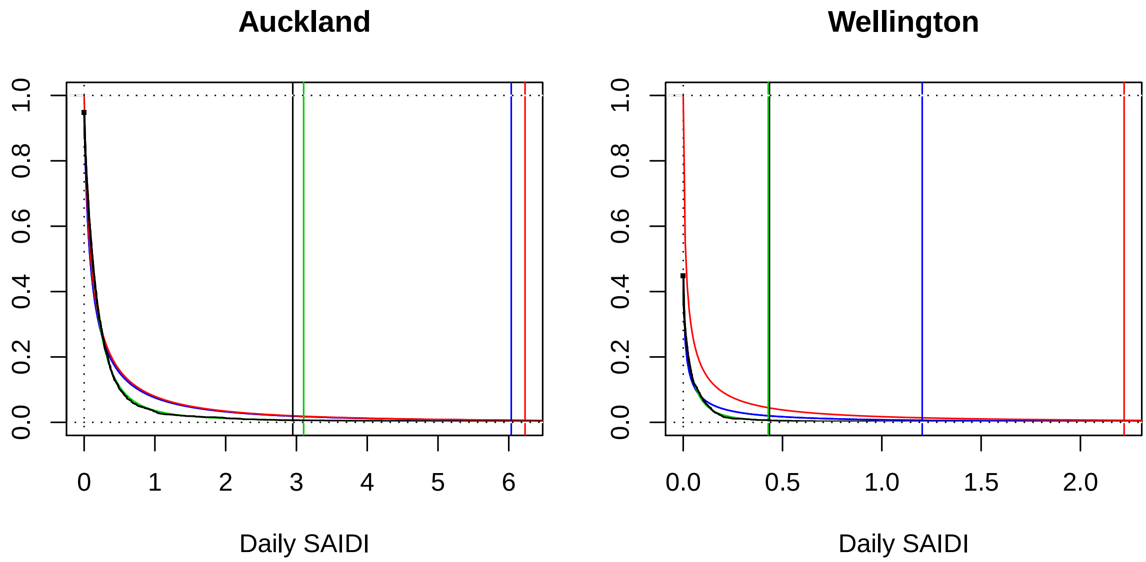

Figure 3: Empirical survivor functions of daily SAIDI values for Auckland and Wellington over the period 1 April 2002 to 31 March 2007 with best fitting survivor functions superimposed. The vertical lines give the corresponding values of T_{MED} for the log-normal with zero SAIDI days excluded (red) or accounted for (blue), the log-normal mixture adjusted for zero SAIDI days (green), and the empirical 99.38% percentile determined directly from the data (black).

of the method proposed in ComCom (2007).

5.3 Recommendations

The shortcomings of the normalisation procedure proposed in ComCom (2007) and discussed in the previous sections are addressed in the following recommendation.

Recommendation 1 *With regard to the normalisation of daily SAIDI and SAIFI values, it is recommended that*

(a) estimates of the threshold T_{MED} *be based on the non-normalised daily SAIDI and SAIFI values in the five year reference period;*

(b) normal variation be assessed using truncated daily data (MEDs excluded) in keeping with IEEE standard 1366^{TM}*-2003;*

(c) the assumption that non-zero daily SAIDI or SAIFI values follow a log-normal distribution be replaced by the assumption that they follow a more flexible family of distributions such as a mixture of log-normal distributions;

(d) zero SAIDI values be accounted for in determining T_{MED}*.*

The above recommendations are straightforward to implement and have the virtue of making the normalisation process simpler, more transparent, and fairer to all New Zealand electricity distribution networks.

6 Dead-bands

The Commission seeks views on appropriate methods for constructing dead-bands within which a quality measure would be allowed to vary. If a network's quality measure falls above the upper limit of its dead-band, then this would constitute a breach of the quality standard and indicate non-compliance. However the Commission does not clarify what will happen if the quality measure falls below the lower limit of the dead-band. Will the network concerned be suitably rewarded and, if so, how? The nature of the implied cost function needs to be made more explicit and, in particular, whether the dead-band should be two-sided or one-sided. In what follows, only a one-sided dead-band and the determination of its upper limit is considered, although the approach adopted can readily be extended to include the two-sided case.

A suitable statistical framework within which dead-band limits can be established is hypothesis testing. Suppose that the normalised daily SAIDI or SAIFI data from the five year reference period follow the theoretical distribution $F(x)$ where the parameters of this distribution are known. In practice $F(x)$ will be estimated from the daily data by a suitable best fitting distribution, such as the log-normal mixture for non-zero SAIDI values with zero SAIDI days accounted for. Like its empirical counterpart, $F(x)$ will also give rise to survivor functions such as those shown in Figure 3. Here it is assumed that $F(x)$ can be regarded as known and not subject to statistical variation. However, if it was felt that the statistical variability of the reference data needed to be accounted for, then the arguments given below can be suitably modified.

Now consider the normalised daily SAIDI or SAIFI data for an assessment year and, in particular, its sample mean $\bar{X}$. Under the hypothesis that the network concerned is compliant (i.e. the assessment year data comes from distribution $F(x)$), one can determine the resulting sampling distribution of $\bar{X}$ and a suitable threshold $\bar{x}_c$ such that the probability

$$\alpha = P(\bar{X} > \bar{x}_c | \text{network is compliant})$$

is suitably small ($\alpha = 0.05$ for example). The constant $\bar{x}_c$ is the dead-band upper limit above which a network would be deemed to be in breach of its quality standard. It is a function of the parameters specifying $F(x)$ (determined from the reference period), and α can be regarded as a *false-positive rate* since it is the probability that a network is in breach of its quality standard when, in fact, it is actually compliant. In the language of hypothesis testing, α is also called the Type I error or size of this test for a breach of the quality standard.

The fairness principle embodied in IEEE standard 1366^{TM}-2003 is readily extended to this situation by requiring that the false-positive rate α is an agreed common probability across all networks. Although the thresholds $\bar{x}_c$ will be network dependent, the same proportion of errors (network deemed to be in breech of its quality standard when it is compliant) will be experienced by each network, on average, over any given period of time.

Under the hypothesis of compliance, the sampling distribution of $\bar{X}$ can be determined in principle, either analytically or by simulation, for normalised samples drawn from the distribution $F(x)$. However, since the sample size for any assessment year is large (365 daily SAIDI or SAIFI values), asymptotic approximations to the distribution of $\bar{X}$ are likely to be sufficiently accurate for the purposes of determining $\bar{x}_c$. Here the distribution of $\bar{X}$ can be shown to have an asymptotic normal distribution with mean μ and standard deviation $\sigma/\sqrt{n}$ where

$$\mu = \mu_{MED}, \qquad \sigma^2 = \sigma^2_{MED}/(1 - p_{MED})$$

when $\bar{X}$ is computed from truncated data, and

$$\mu = (1 - p_{MED})\mu_{MED} + p_{MED}T_{MED}$$
$$\sigma^2 = (1 - p_{MED})\sigma^2_{MED} + p_{MED}(1 - p_{MED})(T_{MED} - \mu_{MED})^2$$

when $\bar{X}$ is computed from censored data. Here $p_{MED} = 0.0062$ is the probability of a MED, and μ_{MED}, σ_{MED} are the mean and standard deviation respectively of a random variable X with distribution $F(x)$ that has been conditioned on X falling below T_{MED}. These conditional means and variances are readily determined from the distribution $F(x)$, either through analytical formulae derived from the statistical model adopted, or directly using suitable empirical estimates from the reference period.

These large sample approximations with $n = 365$ give a dead-band upper limit of

$$\bar{x}_c = \mu + z_\alpha \sigma/\sqrt{365}$$

where $\Phi(z_\alpha) = 1 - \alpha$ and $\Phi(z)$ is the standard normal distribution function. The constant z_α is 1.645 when $\alpha = 0.05$ and, as noted above, μ and σ are calculated from the best fitting distribution of the daily SAIDI or SAIFI data over the five year reference period. This test procedure has the virtue of simplicity, but other more powerful model-based procedures could also be used. This remains a topic for further discussion and research.

These observations lead to the following recommendation.

Recommendation 2 *With regard to dead-bands it is recommended that*

(a) dead-bands be constructed using the principles of statistical hypothesis testing with daily data from the reference period used to establish the hypothesis of compliance;

(b) the false-positive rate α, the probability a network is deemed to be in breach of its quality standard when it is in fact compliant, be set at a common value for all networks in keeping with the fairness principle of IEEE standard 1366^{TM}-2003.

7 Assessing compliance

The Commission is concerned with assessing whether sustained material deterioration in quality standards have occurred over time. If the periods of time envisaged are of the order of two or even three years, then a breach of a quality standard in any one year may not, by itself, provide sufficiently reliable evidence of sustained deterioration over years.

In particular, if α is not small (the Commission's example of $z_\alpha = 1$ yields a false-positive rate of 15.87%), then a relatively large number of compliant networks will be deemed to be in breach of their quality standards. To reduce the occurrence of technical breaches such as these, the Commission have suggested that the assessment of compliance be undertaken using *assessed values* where these are based on the performance of the quality measures over the last three years.

One statistical approach that addresses this issue is *sequential hypothesis testing* where the assessment of compliance at the end of each assessment year would be based on the normalised daily SAIDI or SAIFI data over all available past years in the regulatory period. This approach requires the determination of suitable limits $\bar{x}_c$ that depend on the assessment year concerned. These limits are chosen to ensure that the overall probability of a network being assessed as non-complaint, when it it is compliant, is a very small value and the same for all networks.

Alternatively, an assessment of compliance could be based on the annual assessments of breaches in quality standards, as discussed in the previous section. Perhaps the simplest example of such an approach would be to assess a network as non-compliant at the end of an assessment year if it breached its quality standards in both the assessment year and also the preceding assessment year (*"two in a row"*). In essence the first assessment year resulting in a breach would constitute a warning that a network was in danger of being assessed as non-compliant, with that assessment confirmed in the following assessment year if the network continued to breach its quality standards. If the quality measures of successive assessment years are statistically independent, then the false-positive rate for this simple scheme is α^2. In this case the false-positive rate is 0.25% when $\alpha = 0.05$ and 2.52% when $z_\alpha = 1$. This yields a substantial decrease on the false-positive rate α for individual assessment years. A possible complication with such an approach is the need to separately specify the assessment of non-compliance for the inital assessment years (just the first year in the simple example given above).

The Commission suggests two possible approaches along these lines for improving the reliability of the assessment of non-compliance. It is not entirely clear whether these are to be seen as alternatives, or to be applied collectively.

The first of the Commission's suggested approaches (a so-called three-year moving average) defines a three-year quality measure (assessed value) for an individual network as the smaller of the annual quality measure in the current assessment year, and the average of the annual quality measures in the current and two previous years (three year average). The assessed value is then compared to the network's associated quality standard or reliability target.

For this approach it is noted that the assessed value $\bar{X}_3$ always equals the three-year average when the quality measure for the current year $\bar{X}_1$ exceeds the average $\bar{X}_2$ of the quality measures in the two previous assessment years, and equals $\bar{X}_1$ otherwise. This implies that

$$\bar{X}_3 = \begin{cases} \bar{X}_1 & (\bar{X}_1 \le \bar{X}_2) \\ \frac{1}{3}\bar{X}_1 + \frac{2}{3}\bar{X}_2 & (\bar{X}_1 > \bar{X}_2) \end{cases}$$

and it can be shown that $\bar{X}_3$ is a downwards-biased estimator of the quality standard under the hypothesisis of compliance. Issues associated with this three-year moving average approach are as follows.

Initialisation How $\bar{X}_3$ should be determined in the first two years of the regulatory period is not clear.

Dead-bands It is not clear how the dead-band limits are to be incorporated in this case. The sampling properties of $\bar{X}_3$ under the hypothesis of compliance can be determined, but are not straightforward.

False-positive rate The false-positive rate of this procedure should be the same for all networks (under the hypothesis of compliance), but needs to be evaluated. If, for example, the quality standards are used without allowance for sampling variation (no dead-bands) then the false-positive rate is 35%.

Transparency While superficially simple, the statistical properties and basis of this ad-hoc approach are not so simple and need to be better understood. This is possible, but not necessarily straightforward.

No doubt these issues can be addressed. However it is likely that there are simpler ways of improving transparency and lowering false-positive rates.

The second of the Commission's suggested approaches is conceptually simpler and more in keeping with the "two in a row" approach outlined before. Here a network would be considered in breach of compliance in any assessment year if its assessed values failed to meet their reliability targets in the current year and at least one of the previous two years. In this case the false-positive rates can be evaluated relatively simply and vary by assessment year within the regulatory period. For all bar the first two assessment years, the false-positive rate is approximately $2\alpha^2$ provided α is small. However, the added complication of this approach does not seem justified when compared to the simpler "two in a row" approach. An appropriate choice for α should ensure an acceptable false-positive rate and suitable dead-band upper limits.

These observations lead to the following recommendation.

Recommendation 3 *With regard to assessing compliance over time, it is recommended that*

(a) procedures such as the simple "two in a row" approach be adopted to enhance transparency while retaining flexiblity in setting dead-band upper limits and controlling false-positive rates;

(b) the false-positive rates of any competing approaches be explicitly evaluated and compared.

Finally, in addition to the assessment of normal variation discussed above, there also needs to be a separate assessment of MEDs for compliance.

References

Billinton, R.R. and Acharya, J.R. (2006) Major event day segmentation. *IEEE Transactions on Power Systems* **21**, 1463-1464.

Christie, R.D. (2003a) Statistical classification of major event days in distribution system reliability. *IEEE Transactions on Power Delivery* **18**, 1336-1341.

Christie, R.D. (2003b) Zero SAIDI days issue - response to WMECO. Technical report. http://www.ee.washington.edu/faculty/christie/

Christie, R.D. (2003c) Maximum likelihood estimates for alpha and beta with zero SAIDI days. Technical report. http://www.ee.washington.edu/faculty/christie/

ComCom (2007) *Supplementary guidelines for investigating breaches of the reliability criterion of the quality threshold.* Regulation of Electricity Lines Businesses, Targeted Control Regime. Commerce Commission.

ComComm (2009) *Reset of Default Price-Quality Path for Electricity Distribution Businesses.* Regulation of Electricity Lines Businesses, Targeted Control Regime. Commerce Commission Discussion Paper.

Eto, J.H. and LaCommare, K.H. (2008) Tracking the reliability of the US Electric Power System: an assessment of pubicly available information reported to state public utility commissions. Report commissioned by the US Department of Energy. Ernest Orlando Lawrence Berkeley National Laboratory.

IEEE (2004) IEEE Guide for Electricity Power Distribution Reliability Indices. IEEE Std 1366^{TM}-2003, IEEE Power Engineering Society.

PBA (2007) System reliability: final recommendations for identifying extreme events. Report prepared for the Commerce Commission by Parsons Brinckerhoff Associates.

Warren, C.A. and Saint, R. (2005) IEEE reliability indices standards: major event day calculations and how they relate to small utilities. *IEEE Industry Applications Magazine* **11**, 16-22.