

## Zero SAIDI Days Issue - Response to WMECO

Richard D. Christie  
Department of Electrical Engineering Box 352500  
University of Washington  
Seattle, WA 98195-2500  
christie@ee.washington.edu  
January 24, 2003

### 1. Introduction

I will make a brief response to the WMECO document W.J. Burley and R. Howes, "Review of IEEE 1366-2001 Classification of Major Event Days", Draft 1/24/2003, Western Massachusetts Electric Company.

The WMECO document recommends omitting days with zero SAIDI from the calculation of Alpha and Beta, which are estimates of the parameters of the log normal distribution. The recommendation is based on the following:

1. No "valid statistical rules" justify replacing zero SAIDI days with the minimum SAIDI value, as described in the current proposed guidelines.
2. WMECO's recommended method obtains Major Event Day thresholds that identify a number of major event days per year that WMECO finds satisfactory.

WMECO also considered replacing zero SAIDI days with average SAIDI and median SAIDI.

### 2. Theoretical Discussion

In theory, an ideal log-normal distribution will never generate samples (daily SAIDI values) with values of zero. In practice, zero values appear because the real process is not exactly log normal - it has some discrete components (faults may or may not occur) as well as continuous - and because of the quantization of time (SAIDI per **day**). It may be useful to think of the sampling process as going through a round-off process in which daily SAIDI values below some minimum are rounded to zero.

These theoretical pre-roundoff sample values (there is no way to measure their actual values) are all less than some minimum but greater than zero and can be thought of as occupying the left hand tail of the normal distribution of the logs of the samples. This is shown in Figure 1. The value of the minimum shown is somewhat large to emphasize the content of the tail of the distribution. The top graph on page 9 of the WMECO document shows the actual location of the minimum for WMECO. Ignoring the collection of values at the minimum for a moment, note that the right hand tail is longer than the left hand tail in the WMECO graph, showing the "tail amputation" caused by the rounding.

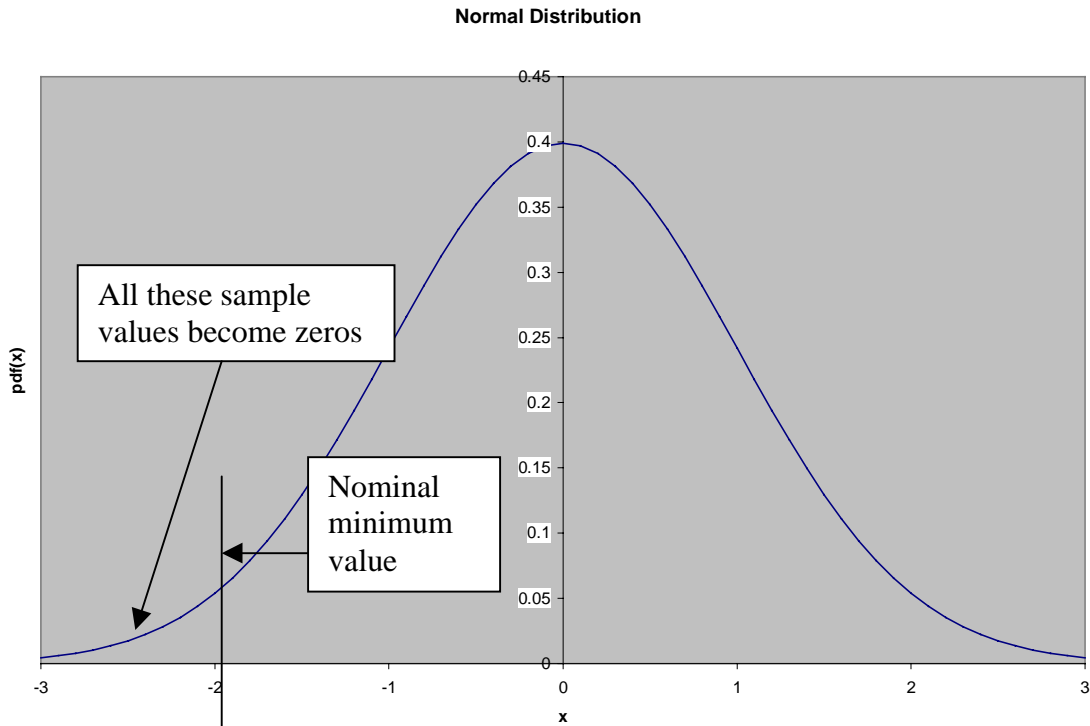


Figure 1 - Normal distribution showing location of zero samples in left hand tail

It is reasonable to guess that the minimum SAIDI value below which actual SAIDI values round to zero is somewhere near the minimum non-zero SAIDI value in the sample set. There is no obvious better approximation, and practical data, for example the first graph on page 9 of the WMECO document, suggest that the approximation is reasonable.

The question, then, is what to do about these zero values, which cannot be used to compute the mean and standard deviation of the logs of the data, (Alpha and Beta, respectively) because their logs are negative infinity.

In answering this question, keep in mind that large utilities often have no days with zero SAIDI values in five years, and calculate the "correct" values of Alpha and Beta using all their daily SAIDI values. For equity among utilities of different sizes, the values of Alpha and Beta computed when zero SAIDI days exist should be as close to the "correct" values as possible.

There are four proposed methods at present. Each will be discussed on its own page with its own figure. (Hence the blank space at the bottom of each page.) The methods are:

- Omitting zero SAIDI days from Alpha and Beta calculations (the WMECO proposal)
- Replacing zero SAIDI days with the minimum SAIDI value (the current proposed standard)

- Replacing zero SAIDI days with the average SAIDI value
- Replacing zero SAIDI days with the median SAIDI value

A few other methods are briefly described later. I have not had time to research them in detail.

### 3. Effect of Omitting Zero SAIDI Days

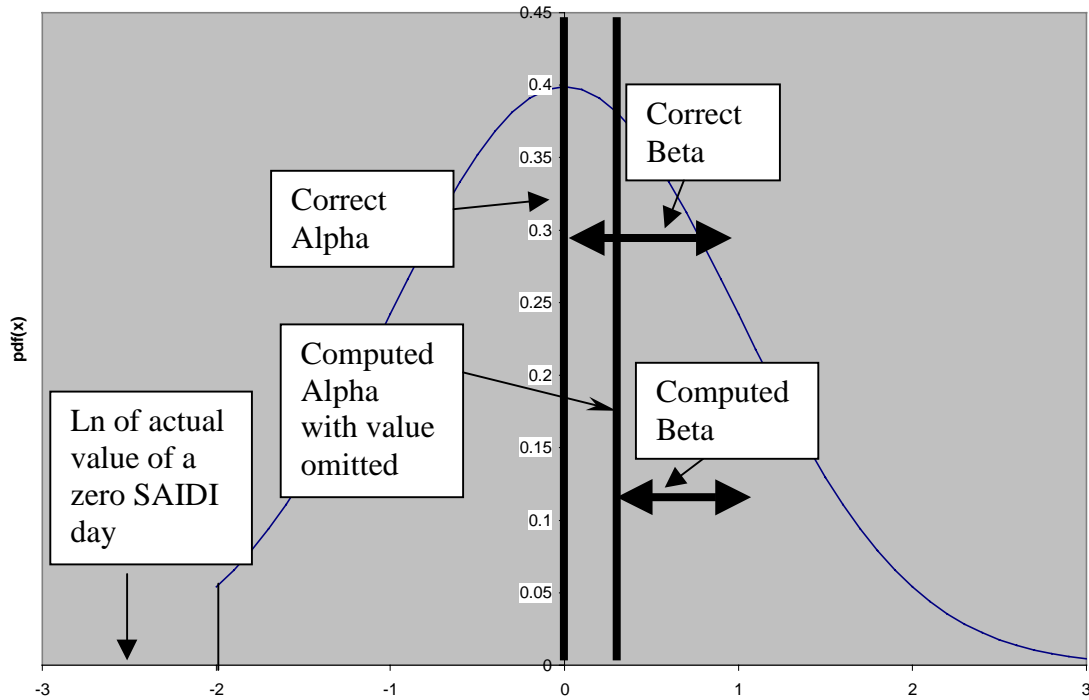


Figure 2 - Graphical illustration of effects of ignoring zero SAIDI days in calculating Alpha and Beta. Effects are qualitative, not computed.

The first option is to ignore the zero values. This means that Alpha and Beta are calculated from a normal distribution with the left hand tail chopped off. Consider what happens when one zero SAIDI value is ignored. The mean (Alpha) get larger than the actual value and the standard deviation (Beta) gets smaller. The effect on major event day threshold is unclear, but probably results in an increase compared to the actual value. This is illustrated conceptually in Figure 2. I have not had time to do a numerical example. However, the qualitative effect - the increase in Alpha and the decrease in Beta - is clear from statistical theory.

#### 4. Effect of Replacing Zero SAIDI Days with Minimum SAIDI Value

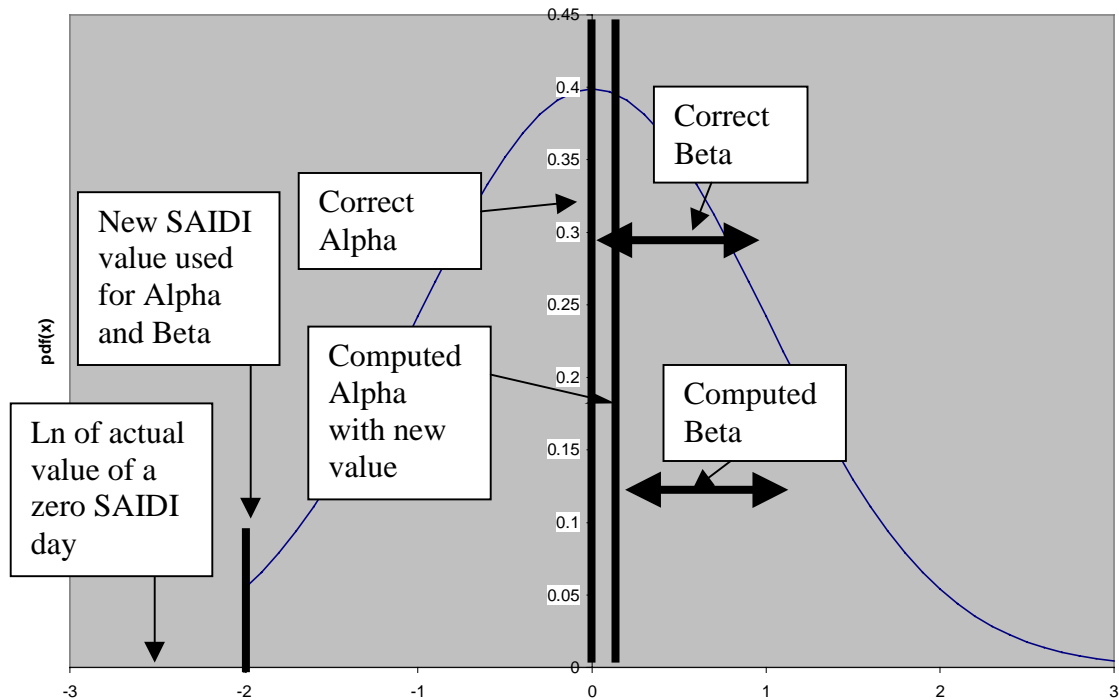


Figure 3 - Graphical illustration of effects of replacing a zero SAIDI day with minimum non-zero SAIDI on calculating Alpha and Beta. Effects are qualitative, not computed.

The second option is to replace the zero SAIDI values with the minimum SAIDI value. This results in the distribution shown in the top figure of page 9 of the WMECO report, and is illustrated conceptually in Figure 3.

Consider the effect on Alpha. Because the SAIDI value is moved to the right (increased), computed Alpha will be larger than the actual. But because a value has been added that is below the actual Alpha value, the computed value will be less than if the zero SAIDI point was omitted. Thus the computed value of Alpha is closer to the actual value than if zero SAIDI points are omitted.

Similarly, the computed Beta value will be less than the actual Beta value but larger than if the zero SAIDI point is omitted, and thus closer to the actual Beta value.

Because the Alpha and Beta values calculated by replacing zero SAIDI days with minimum SAIDI value are closer to the actual values than if the zero SAIDI days are omitted, this method introduces less error and more accurate results than omission, and should be used instead of omission.

## 6. Effect of Replacing Zero SAIDI Days with Average SAIDI Value

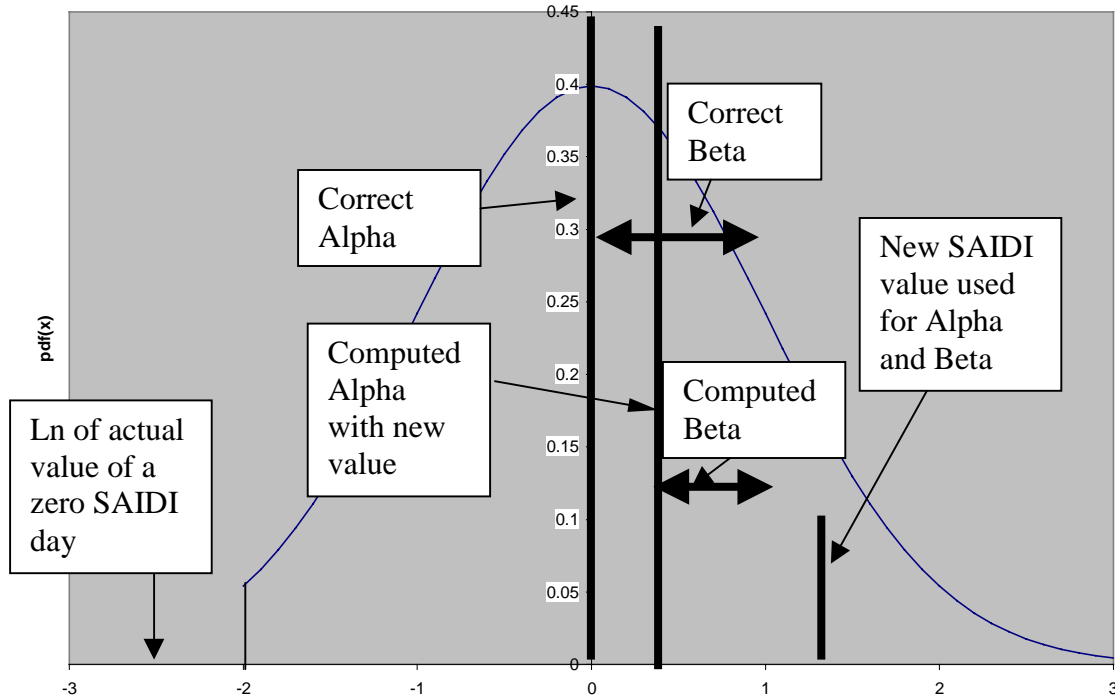


Figure 4 - Graphical illustration of effects of replacing a zero SAIDI day with the average SAIDI value on calculating Alpha and Beta. Effects are qualitative, not computed.

The third option is to replace zero SAIDI days with the average SAIDI value calculated from the actual daily SAIDI values. It is characteristic of the log normal distribution that the log of the mean (average) of the samples will be higher than the mean of the logs, that is, the replacement value will be higher than the actual Alpha. This effect is shown in Figure 4, and also in the figure on page 6 of the WMECO document.

Consider the effect on the computation of Alpha. Because the replacement value is larger than the actual value, and the difference between the actual value and the replacement value is far larger, the mean (Alpha) will be much larger, and farther from the correct mean, than if the minimum value is used. The mean is also larger than if the zero SAIDI day were omitted. (Because the replacement value is larger than the omitted mean, including it would increase the mean further.)

The effect on standard deviation (Beta) is less clear. By eye, the average value is closer to its computed mean than the minimum value, so standard deviation should decrease compared to that case, and thus be less accurate. However, standard deviation would appear to increase compared to the omission case.

Since both Alpha and Beta of the average replacement case are less accurate than the minimum replacement case, the minimum replacement case is theoretically more accurate and should be preferred.

## 7. Effect of Replacing Zero SAIDI Days with Median SAIDI Value

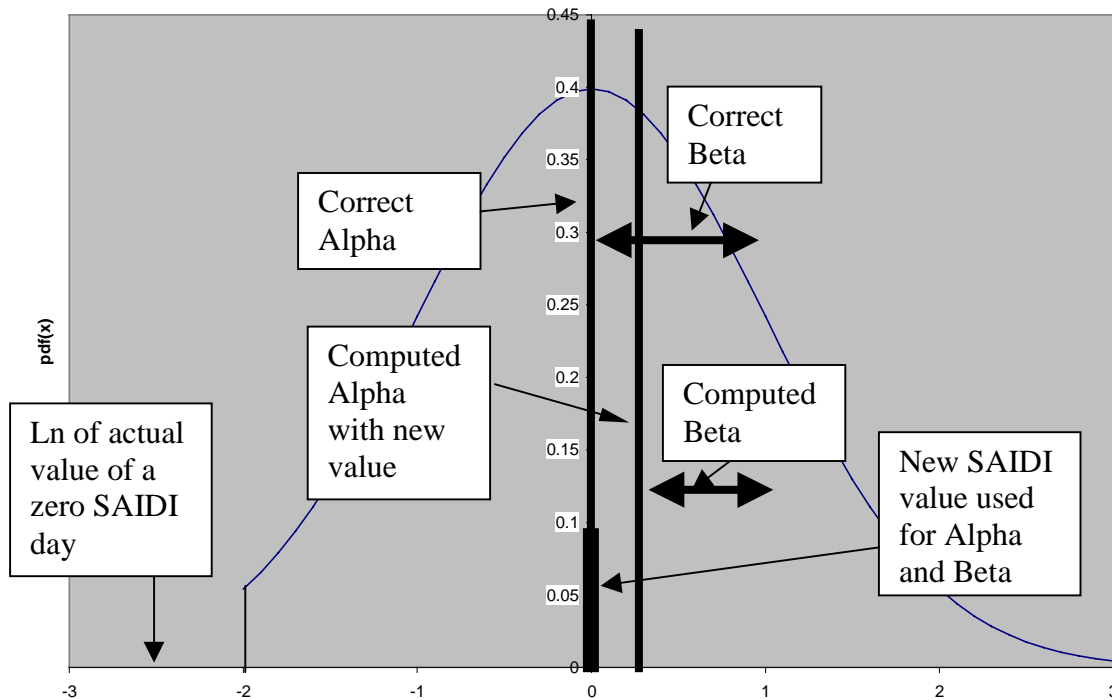


Figure 5 - Graphical illustration of effects of replacing a zero SAIDI day with the median SAIDI value on calculating Alpha and Beta. Effects are qualitative, not computed.

The last option is the use of the median value of SAIDI as a replacement value for zero SAIDI days. This is the method currently proposed by WMECO. The log of the median value of SAIDI is the mean (average) of the distribution of the logs of the values, i.e. in theory it is Alpha. This can be seen in the figure on page 7 of the WMECO report, and in Figure 5.

Because the replacement value is larger than the minimum computed Alpha is larger than the actual Alpha. Because the log of the median is farther from the actual sample value than the log of the minimum, the effect on Alpha is larger, and computed Alpha is further from actual Alpha.

Because the replacement value is close to the computed mean, it makes little contribution to standard deviation, and the computed standard deviation (Beta) will be less than the actual value, and less than the value computed using the minimum SAIDI as a replacement.

Using the median as a replacement value provides values for Alpha and Beta that are better than using the average, but still creates more error than using the minimum value.

## 8. Summary of Analysis



Of the four options for dealing with zero SAIDI days, theoretical considerations show that use of the minimum SAIDI value introduces the least error into the calculation of Alpha and Beta.

This conclusion is contrary to the assertions of the WMECO document, which are not supported by theoretical analysis. The minimum SAIDI value is not a fabricated value. It is an approximation of the actual theoretical value of the zero SAIDI day, and has been shown to be the best approximation among the four alternatives considered using valid statistical rules. Its use does in fact enhance the overall accuracy and consistency of the method. Use of the minimum value does not "skew the  $T_{MED}$  against the Utility" (WMECO p. 3). Rather, it more closely approximates the actual values of Alpha and Beta, permitting a more equitable computation for  $T_{MED}$  among large and small utilities.

## 9. Other Methods of Estimating Alpha and Beta

There are other methods available for estimating the values of Alpha and Beta that are not affected by zero SAIDI days. Note that the actual values of Alpha and Beta can never be calculated. All of the methods discussion for calculating Alpha and Beta are more correctly described as methods for estimating the value of Alpha and Beta, and every estimate will have errors.

These methods are:

Calculation of Alpha and Beta from the mean ( $\mu$ ) and standard deviation ( $\sigma$ ) of the reliability values. The relevant formulae are

$$\beta^2 = \ln\left(1 + \frac{\sigma^2}{\mu^2}\right) \quad (1)$$

and

$$\alpha = \ln(\mu) - \frac{\beta^2}{2} \quad (2)$$

Note that using these formulae result in values of Alpha and Beta significantly different from those obtained by the current proposed method of taking the average and standard deviation of the logs of the daily SAIDI values, even for data sets with no zero days. This is because the different estimation methods experience different effects from randomness in the data. The proposed method is preferred to this one because it is closer to the Maximum Likelihood Estimates (MLE) of Alpha and Beta. See Limpert, Stahel and Abbt, "Log-normal Distributions across the Sciences: Keys and Clues", *BioScience*, May 2001, Vol. 51, No. 5, pp. 341-352, on the web at <http://stat.ethz.ch/~stahel/lognormal/bioscience.pdf>. This is also a good general discussion of log normal distributions.

Another method briefly described in Limpert *et. al.*, is the use of quartiles as a robust estimator Beta. There has not been time to learn more about this, but again the MLE

estimator is probably preferable. Also, quartiles may be vulnerable to zero days if more than a quarter of the days have zero SAIDI values. This is possible for small utilities.

There may be other methods in the literature that I have not had time to explore completely.

## **10. Conclusion**

This discussion has shown that the use of minimum SAIDI values to replace zero SAIDI values for computation of Alpha and Beta provides the least error in the computation, and hence the least error in  $T_{MED}$  and number of identified major event days.

WMECO's concerns about the number of major event days identified using the method are similar to the concerns a number of different utilities have voiced over every proposed method of identifying them - that the number of days identified does not seem right to the utility. In this case, theory shows that the method proposed by WMECO results in more - not less - error in the classification process, even though the number of identified days increases to a level satisfactory to the utility. It may well be that WMECO actually did not have any major events in the past three years. It would be natural to recall the largest event WMECO experienced in the past three years and identify it as a major event - but it might well not be a major event when considered over a longer time frame.

The preference of utilities to see at least some major events identified over any given time period is understandable. The previously proposed bootstrap method was intended to address exactly this problem by ensuring that a fixed number of MEDs were identified in a multi-year data set. If none of them happened to be in the current year, at least utilities would see that MEDs were being identified in some years. However, this method was rejected when tested by several utilities because they did not like the results obtained.