Spectrum Planning

— Analysis of methods for the summation of Log-normal distributions

Karina Beeke
National Grid Wireless

When carrying out coverage predictions for RF signals, statistics play a big part and the statistical nature of the predicted values cannot be ignored. In the particular case of location variation, the signals are assumed to follow a log-normal distribution and various methods are available for carrying out summations of such signals.

This article examines the different algorithms in an attempt to assess the suitability of each one and to identify the optimum method to use. Two main scenarios are considered. The first looks at the summation of a series of signals with various mean values, such as might be used when summing the contributions of a number of interferers. The second looks at the best method of including a constant such as the minimum field strength. In all cases, the impact of the mean level and standard deviation of the contributors is considered.

When carrying out coverage predictions for RF signals, statistics play a big part and the statistical nature of the predicted values cannot be ignored. Consider the variation of signal with position. Suppose the mapping data used has a resolution of 50 m. We may make predictions to points at 50 m intervals; however, in practice, we cannot expect the signal levels to be constant across the whole of the 50 m square. For example, at one point we may be in front of a building but at another point, in the same square, we may be behind it. As a result of this, any prediction signal can be quoted as having a mean value and an associated variance or standard deviation [1]. By use of these values, we can determine whether or not to expect say 99% of locations within a particular square to receive acceptable coverage.

Generally, when considering this location variation, the values are considered to follow a log-normal distribution. This means that the logarithm of the signal level follows a normal, or Gaussian, distribution. Conveniently, the Gaussian distribution is well documented and there are many algorithms for relevant functions. Such a distribution must be taken into account when carrying out the summation of the signals in order to determine whether or not a particular location is served.

To make this clearer, consider a simplified example, together with *Fig. 1*.

> Suppose a signal has a mean value of 68 dBµVm and a standard deviation of 5 dB. Now suppose that we need a signal level of 60 dBµVm in order for a location to be served. Initially, it may appear that coverage is achieved. However, we must look at this more carefully.

1. The variance is the square of the standard deviation and tells us about the spread of the data. Thus a distribution with a small standard deviations has most values clustered around the mean value; a distribution with a large standard deviation is more spread out.

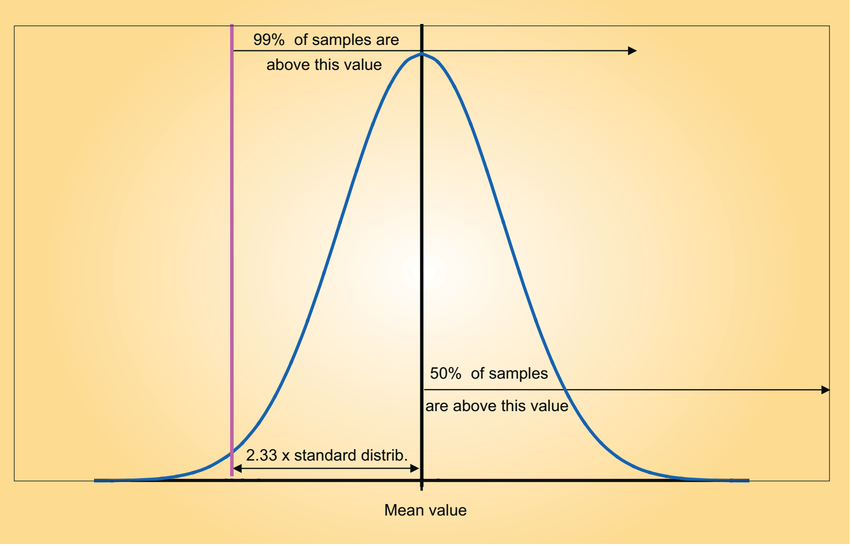

Figure 1
Normal / Gaussian Distribution
(For a log-normal distribution, the abscissa scale must be logarithmic)

The actual signal is (68 – 60) = 8 dB above the required level.

Now, for some services, in particular mobile services, we may decide that acceptable coverage is only obtained if, say, 99% of locations are covered. This is where knowledge of the particular statistical distribution is required. In general, 50% of samples, should have a value greater than the mean and 50% below the mean. However, if we are interested in percentages other than 50%, then we need to know the relationship between that and the standard deviation. For the Gaussian distribution, 99% of samples should have a value of *(mean) – (2.33 x standard deviation)*. Thus, in our case, 99% of locations will have a field strength of 68 – 2.33 x 5 = 56.4 dB. Since this is less than the required 60 dB, we may decide that acceptable coverage will not be achieved.

Let's also see what percentage of locations should lie above the required level of 60 dB.

With a standard deviation of 5 dB, then a difference of 8 dB is 8/5 = 1.6 x standard deviation above the mean. By looking at the inverse cumulative probability function for the Gaussian distribution, we find that this represents just under 95%: i.e. 95% of locations can be expected to have values greater than the *mean – (1.6 x standard deviation)*.

The above example shows why it is so important to get an accurate value for the standard deviation as well as the mean. This is the case, particularly when we are interested in the tails of the distribution, e.g. above 90% or below 10%. Overestimating the standard deviation will result in predictions being pessimistic and the resulting network will be more expensive than it needs to be. Conversely, underestimating the standard deviation may result in too few sites being built and a network that does not perform adequately.

This article examines different algorithms for carrying out summations in an attempt to assess the suitability of each one and to identify the optimum method to use.

The summation of signals is of particular significance in a single frequency network (SFN) where it is necessary to sum the wanted signals, to sum the interfering signals and also to take account of the minimum field strength required to overcome system noise. It is beyond the scope of the article to provide a detailed description of each method. However, a very good overview of several of the methods is given in the EBU document, BPN-066: *"Guide on SFN Frequency Planning and Network Implementation with regard to T-DAB and DVB-T"* which is available to EBU Members only on the EBU's website ... or by requesting a copy from **spectrum@ebu.ch**. Another good overview is freely available on the ITU's website: **http://www.itu.int/ITU-D/pdf/3888-21.3-en.pdf**.

The methods considered are:

- S.C. Schwartz and Y.S. Yeh: **On the Distribution and Moments of Power Sums with log-normal Components**
  BSTJ , September 1982
- A. Safak: **Statistical Analysis of the Power Sum of Multiple Correlated log-normal Components**
  IEEE Transactions on Vehicular Technology, Vol. 42, No. 1, February 1993.

Abbreviations

DAB	Digital Audio Broadcasting (Eureka-147) http://www.worlddab.org/
DVB	Digital Video Broadcasting http://www.dvb.org/
DVB-T	DVB - Terrestrial
LNM	Log-Normal Method
RF	Radio-Frequency
SFN	Single-Frequency Network
T-DAB	Terrestrial - DAB

This is an extension of the Schwartz and Yeh method. For the purpose of this article, the only difference is the calculation of the intermediate functions G1, G2 and G3. Schwartz and Yeh use a polynomial approximation to determine these values whereas Safak uses analytical expressions.

- L.F Fenton: **The Sum of log-normal Probability Distributions in Scatter Transmission Systems** IRE Transactions on Communications Systems, March 1960.

  In general, Fenton considers equivalent log-normal distributions based on the first moment (the mean value) and the second central moment (the variance); the second and third central moments or third and fourth central moments. This current study considers only the first of these which is equivalent to the log-normal method (LNM), as described in BPN-066.

- k-LNM: this method is very similar to LNM but uses a correction factor k; in this study, three values for k have been used: 0.3, 0.5 and 0.7 (denoted by KLMN-3, KLNM-5 and KLNM-7 respectively). **NB:** If k is set to 1.0, then the method is identical to LNM.
- t-LNM v2: This is another variant of LNM, also described in BPN066. To quote from this restricted EBU document, "*It approximates the distribution of the logarithmic sum field strength by a Gaussian distribution which possesses the same mean value and the same variance as the true distribution*". In this study, version 2 has been used which uses a computationally more efficient algorithm.

The results of these various methods have been compared with the results using a Monte-Carlo method [2].

When considering the coverage of a single-frequency network, the analysis may be carried out as follows[3]:

1) Determine the signals contributing to the wanted input and sum [4]. Designate this as the **SumWants**: SW.
2) Determine the signals causing interference to the wanted input and sum [4]. Designate this as the **SumInts**: SI.

   NB: In practice, we must also take into account the nature of the interfering field strengths and apply factors (protection ratios) which are dependent on the system, the relative frequency etc. However, for this study, such factors are not required as it is simply the accuracy of the summation which is of interest.
3) Of course, even if there were no interfering signals, reception may still not be possible because of environmental and system noise. Therefore, we must also determine the minimum field strength (MinFS) required (to take account of system noise) and add this to SumInts 4 (ΣI + MinFS).

---

2. In this method, we generated random values with the appropriate mean and standard deviation, and then added them together. By doing this numerous times, we could then determine the resulting mean and standard deviation of the summed terms. These, then, were the values used to determine the errors of the various methods under investigation.
3. It is beyond the scope of this article to discuss the determination of signal levels contributing to the various terms, inclusion of protection ratios etc. The analysis assumes that all of the appropriate weightings and protection ratios have been included.
4. For all the summations, we are summing powers rather than field strengths; this is implicit in all of the methods used.

4) Finally we can use the above results to determine the mean and standard deviation of $\frac{\sum W}{\sum I + MinFS}$.

   From this we can gain an idea of the way the signal varies and hence calculate the percentage locations covered.

Therefore, the assessment of the summation methods has been split into two parts:

- Part 1: The best method for determining ΣW and ΣI;
- Part 2: The best method for determining ΣI + MinFS and $\frac{\sum W}{\sum I + MinFS}$.

Each of these problems will be discussed separately.

Part 1: Summation of log-normally distributed signals

Calculations carried out

Often, it appears that analyses of summation routines use equal mean values when assessing the accuracy of an algorithm. In practice, of course, this is rarely the case. A series of signals arriving from different transmitters will almost always have different means. However, the reception environment may determine the standard deviation of location variation attributed to each signal. Thus one value of standard deviation may be used for a dense urban environment, while a different value is used in rural areas. Nevertheless, for a specific setting, it is often assumed that each incoming signal has the same standard deviation [5]. Thus, in this analysis four sets of data were used:

- A series of signal values obtained from a planning prediction exercise was used as the input values. There were 48 values with varying mean values over a range of 23 dB. Denote this the “Realistic Data Set”.
- A series of regularly-spaced data: 48 values with means varying over a range of 10 dB. Denote this the “10 dB Data Set”.
- A series of regularly-spaced data: 48 values with means varying over a range of 1 dB. Denote this the “1 dB Data Set”.
- A series of data with constant means of 0 dB. Denote this the “0 dB Data Set”.

In each case, the data was sorted and the largest terms summed first.

The first set of data was assumed to be realistic in terms of the type of summation needed in practice. This was simply a set of results picked at random and happened to have 48 samples. The other three sets of data were then used to assess the sensitivity of the methods. This was considered necessary in case the “realistic” data produced atypical results.

For each set of data, the summations were repeated for different standard deviations, varying between 4 and 8 dB. Note that, for each particular summation, it was assumed that all 48 contributing terms had the same standard deviation.

The following terms were calculated:

- resultant Inverse cumulative probability function;
- resultant mean;
- resultant standard deviation.

Generally, it is assumed that the sum of two log-normal variables may also be approximated by another log-normal distribution. This is the reason for including the first bullet point in the calcula-

5. Denote this the “Input” standard deviation

tions. Comparison of these values with the corresponding values for a true log-normal distribution will give an idea of the validity of the assumptions [6].

The results were compared with answers obtained using the Monte-Carlo method with 1 000 000 samples.

Results

Resultant Inverse cumulative probability function

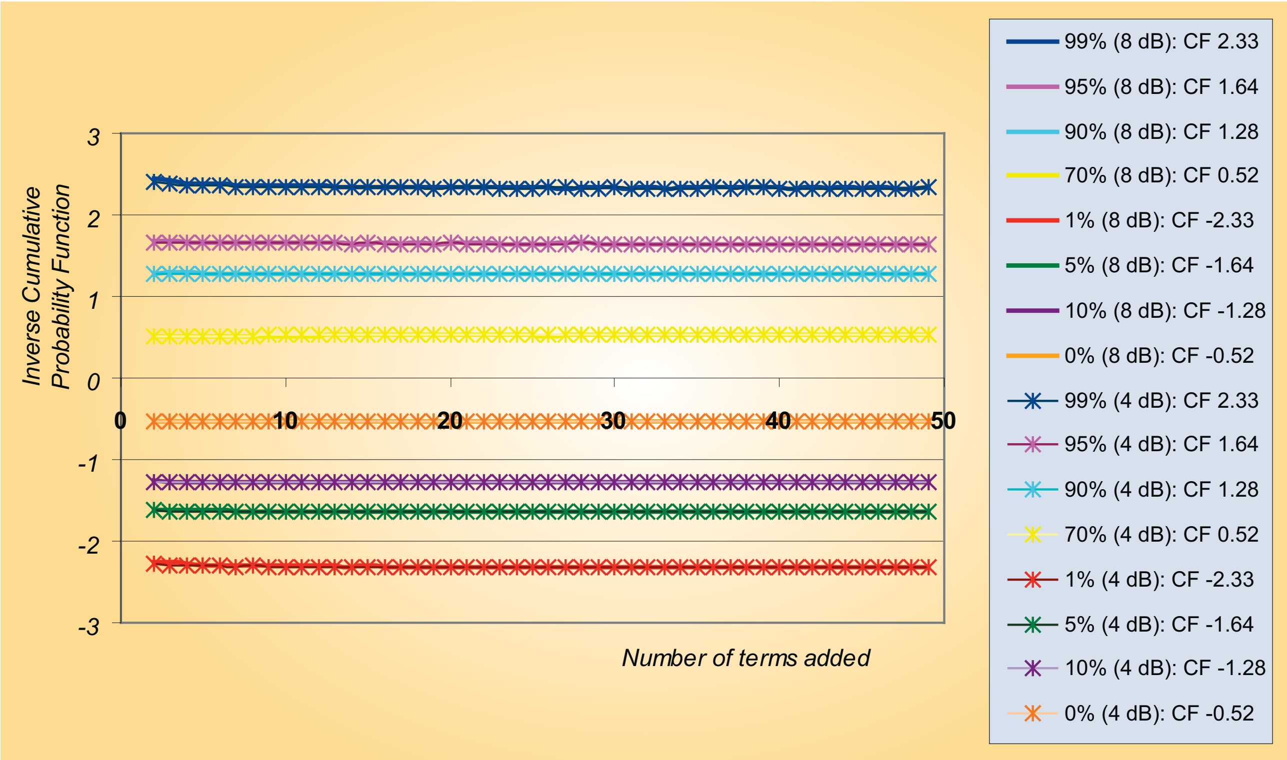

Figure 2
Inverse cumulative probability function

Fig 2 shows the resultant inverse cumulative probability function for the first set of data ("Realistic Data"). These values were obtained from the Monte-Carlo simulations. Two values of input standard deviation have been used: 4 dB and 8 dB. The results for the other data sets were similar. This suggests that we can indeed treat the resultant summation as following a log-normal distribution, even when many terms are added together.

Resultant mean

Fig 3a illustrates the error in mean value obtained with the different summation methods. For this particular case, the realistic data set was used with an input standard deviation of 5 dB. Unfortunately, it is not possible to include all the graphs in this article. However, as we might expect, there is no single "Best" method; the method that produces the lowest error changes with changing input standard deviation value as well as depending on the number of terms summed together.

6. For example, consider a log-normal distribution with mean μ dB and standard deviation σ dB. Then 99% of the terms will lie below $\mu + 2.33\sigma$. If we find that, for our given summation, the inverse cumulative probability function is not 2.33 at 99%, then we can see that we cannot approximate the resultant by another log-normal distribution.

In spite of this, it is not necessarily useful to apply the method giving the lowest error. In some cases, use of a different method may produce an error which is only very slightly higher.

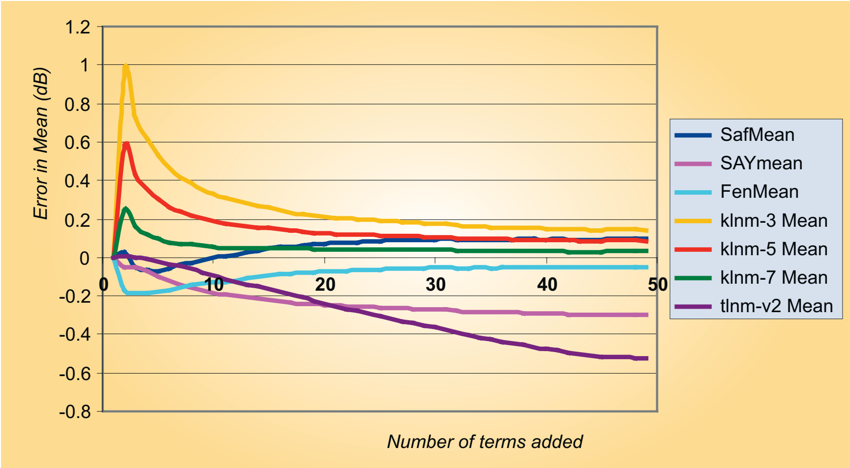

Figure 3a
Error in Means (Realistic Data Set)
(Based on Standard Deviation of 5 dB)

This can simplify things enormously. *Fig 3b* shows an initial and a simplified regime for selecting the method giving the lowest error in mean value. Note, that this is only one possible solution to the problem of balancing implementation complexity and calculation accuracy.

Each filled circle has a colour denoting the summation method to use. The ordinate axis represents the input standard deviation (dB) and the abscissa denotes the number of terms added in the summation. The top chart shows the initial results for the realistic data and the lower chart shows a simplified regime. This latter chart was derived looking at the results of all four sets of data; the resulting errors in the mean value are always less than 0.25 dB for the summations carried out; for all but the highest standard deviation, the errors were always below 0.1 dB.

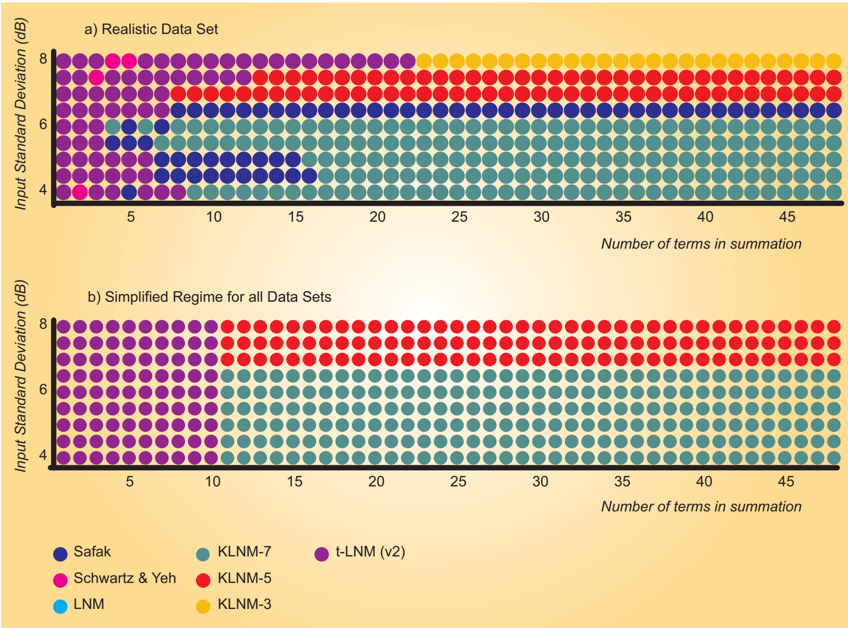

Figure 3b
Minimising the error in Mean

Taking this into account suggests that, for the first ten terms, we could use the t-LNM (v2) method; thereafter use the k-LNM method with k=0.5 or 0.7 depending on the input standard deviation.

Resultant standard deviation

Interestingly, the method that gives the lowest error in mean value is not necessarily the same as the method resulting in the lowest error in standard deviation. An example is shown in *Fig 4*. Comparison of *Figs 3a* and *4* clearly demonstrates this.

In most cases, method t-LNM (v2) results in the lowest error in resultant standard deviation. Moreover, where the t-LNM (v2) method does not give the lowest error, the difference from the minimum error is very small.

However, it seems appropriate to recommend t-LNM (v2) as the method to use when determining the standard deviation of the sum of log-normal variables.

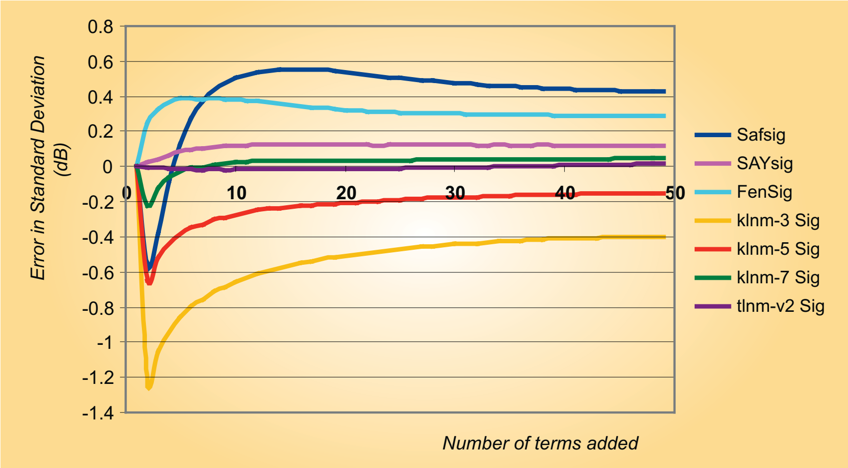

Figure 4
Error in Standard Deviation (Realistic Data Set)
(Based on Standard Deviation of 5 dB)

Discussion

In summary, the results of this part of the study indicate that, when we sum together a series of log-normal variables, we can consider the resultant to be approximated adequately by another log-normal distribution.

As we become interested in values towards the tail of the distribution, such as the values representing 99% of locations, then the accuracy of standard deviation becomes increasingly important.

Computer speeds are ever increasing. In view of this, when more than ten terms are to be added, there would appear to be advantages in carrying out two calculations; the first to determine the mean and the second to determine the standard deviation.

Part 2: Sum of log-normal distributions: addition of a constant

In the introduction to this article, we looked at the four steps which may be needed when carrying out coverage analysis. Part 1 of this document considered steps (i) and (ii), determining the sum of the wanted signals (SumWants – ΣW) and the sum of the interferers (SumInts – ΣI).

In part 2 it is assumed that the means and standard deviations of SumWants and SumInts have already been determined using the most appropriate method. The question now is, "What is the best way to take into account the Minimum Field Strength (MinFS) to enable the final resulting value of percentage locations served to be predicted accurately?"

Therefore, each of the methods described previously was used with the following:

1) SumInts – standard deviation: varies from 0.5 to 8 dB.

   NB: This is the standard deviation of the *sum*, not the standard deviation of the individual interferers. In general, the resultant standard deviation will be lower than the input standard deviation of the individual terms.

2) SumInts – mean: varies from –25 dB to 25 dB relative to MinFS.

3) SumWants – standard deviations: varies from 0.5 to 8 dB
   Again, this is the standard deviation of the Sum.

4) SumWants – mean varies from 0 to 20 dB above the mean of SumInts or the mean of MinFS (whichever is the greater)

Note, in this case, another method was also used, not mentioned previously: This is denoted by "EBUalt", and is also described in BPN-066.

In this method, the percentage of locations covered is determined as follows:

- Determine the percentage of locations covered, taking into account the wanted and interferers only. (ie neglect MinFS). Express as a fraction.
- Determine the percentage of locations covered, taking into account the wanted and MinFS only. (ie neglect the interferers). Express as a fraction.
- Find the product of these.

For example, if taking account of the interferers only gives 90% of locations served and taking account of MinFS only results in 80% of locations served, then the overall percentage of locations covered is taken to be: 0.90 x 0.8 x 100 = 72%.

Having carried out this summation, the resulting means and standard deviations were used to determine the overall mean and standard deviation of $\frac{\sum W}{\sum I + MinFS}$ and hence a value for the percentage locations covered was obtained [7]. This was compared with the value of percentage locations served, obtained via the MonteCarlo method. In this case, the Monte-Carlo method used 10^5 terms.

The next stage was to analyse the results and determine a method of finding the best rule for carrying out the summation. *Fig 5* is just one of many produced and shows which method gives the lowest error for a given combination of means and standard deviations.

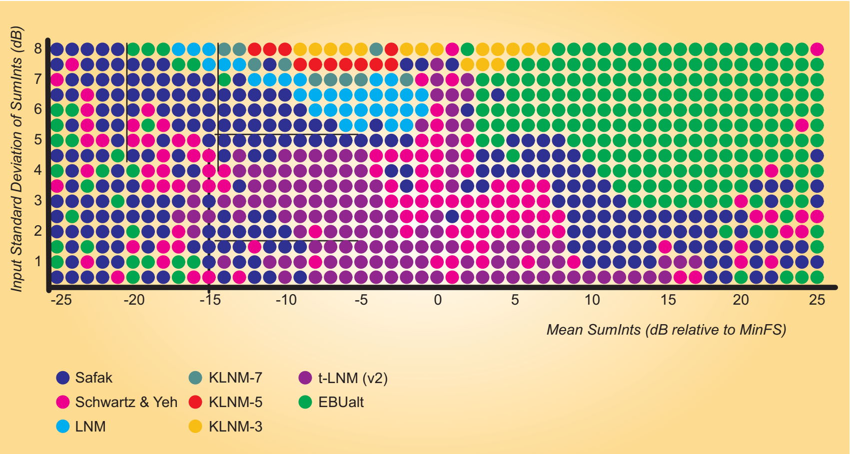

Figure 5
Chart depicting optimum method to include MinFS when determining percentage locations covered

The abscissa indicates the mean of SumInts in dB above MinFS. The ordinate axis represents the standard deviation of SumInts and varies from 0.5 to 8 dB. The calculations were carried out in 1 dB steps for the SumInts mean and 0.5 dB steps for the SumInts standard deviation.

In this particular case, the mean of SumWants is 10 dB. i.e. 10 dB above SumInts on the right-hand side of the chart where mean SumInts is positive, and 10 dB above MinFS on the left-hand side where mean SumInts is negative. Furthermore, the standard deviation of SumWants was taken to be 3 dB.

The figure should be interpreted as follows: Consider the point with coordinates (10, 4). This corresponds to the case where the mean of SumInts is 10 dB above MinFS and the standard deviation of SumInts is 4 dB; for SumWants, the mean is 20 dB above MinFS (10 dB above SumInts) and the standard deviation is 3 dB. The small circle at this point is blue. Thus, for this combination of means and standard deviations, Safak's method will give the lowest error in determining the percentage locations served.

Initially, the next step planned was to produce a simplified regime as was achieved in Part 1. It was considered totally impractical to specify an implementation scheme whereby the summation method changes rapidly with varying means and standard deviations. Therefore, changes to the summation methods were considered to try and optimize the situation with regard to balancing error and simplicity of implementation. For higher values of the mean of SumWants (above 10 dB), this was possible. However, for lower means, this was not the case.

Finally, it was decided that the best approach may be simply to construct a look-up table for the four parameters in order to determine the most appropriate method. This would have an additional advantage that information of the associated error would also be available.

7. This was carried out assuming another log-normal distribution. Thus, for example, it was assumed that, if 99% of locations were served, then the relevant value is at 2.326 x standard deviation away from the mean.

Karina Beeke is a Senior Technologist within National Grid Wireless in the UK and has over 20 years of experience in the broadcast business. Her work at the company focuses on various facets of electro-magnetic theory relating to broadcasting and telecommunications networks; this includes the computational aspects of spectrum planning for both analogue and digital networks from LF to SHF. In addition, she is significantly involved in the analysis of RF Exposure.

Ms Beeke read Engineering Science at the University of Oxford, graduating in 1983. Following this, she worked for the BBC in its Engineering Research and latterly its Transmission departments.

Karina Beeke has participated in the work of the EBU for 15 years, including attendance at CENELEC meetings as an EBU representative. Currently she is the Project Manager of the EBU project groups B/EIC and B/EES.

Conclusions

A range of methods has been tested for carrying out summations with log-normal variables. Two categories of calculation have been investigated:

1) *Summation of a series of log-normal variables.*
Here it was determined that the following approach should be used in order to minimize errors: For summation of 10 or fewer terms, the t-LNM (v2) method should be used; for more than 10 terms, the t-LNM (v2) should be used to determine the standard deviation and the k-LMN method should be used to determine the mean. For input standard deviations up to 7 dB, this results in an error of less than 0.1 dB; for input standard deviations up to 8 dB, this results in an error of less than 0.25 dB.

2) *Combination with a constant*
When adding a constant to a log-normal variable, determining a simple rule to choose the best summation method is much less clear. After much analysis, it is recommended to use a "look-up" table to specify which method to use for various combinations of standard deviations and means. Details of the possible errors could also be associated with such a "look-up" table.