HOW ACCURATE IS MEAN Q – 5 SIGMA DURING SEGMENT COMMISSIONING?

Michael Cuddington

Michael.cuddington@bt.com

BT Global Services, BT, pp TH22, Telephone House, Charter Row, Sheffield, S1 1BA, UK

Abstract: This paper looks at the $\overline{Q}-5\sigma$ figures derived from real data over a couple of years and compares them to minimum figures obtained subsequently, to try and determine whether the use of $\overline{Q}-5\sigma$ is accurate or otherwise.

The paper considers the variation of Confidence Trial length as a function of accuracy.

It is concluded that the accuracy of the $\overline{Q}-5\sigma$ Confidence Trial is not particularly good, for a number of examples, and that is likely to be due to their non-normal distribution.

1 INTRODUCTION

As part of the commissioning of a new system, or one that is upgraded with additional wavelengths, it is standard to monitor the received performance of individual wavelengths for a period of time (usually referred to as a Confidence Trial or Stability Period). For the duration of the Confidence Trial, Q data – derived from the Bit Error Rate of the section – is logged every 15 minutes. At the end of the Confidence Trial, the raw data can be used to determine the average Q and also the standard deviation (sigma). These two figures can then be combined to form a Q Mean – 5*Sigma ($\overline{Q}-5\sigma$) value. This value is then compared to the Commissioning Limit, a contractual figure that must be met by the Supplier.

Whilst, due to margins and performance, the $\overline{Q}-5\sigma$ is usually well above the commissioning limit, this paper looks at whether the result of the $\overline{Q}-5\sigma$ calculation can be used to reliably foretell the worst performance expected over a limited time period in the future.

This paper was written with a desire to see how accurate the calculations are and whether any useful conclusions can be drawn with respect to Confidence Trial length. It is largely an empirical approach, using real historical data to draw conclusions.

Although a lot of emphasis is placed on the importance of analysing Confidence Trial results to ensure that $\overline{Q}-5\sigma$ calculations are compliant against the power budget, it was not clear whether the actual longer term distributions were close enough to Gaussian (Normal) to allow the $\overline{Q}-5\sigma$ calculation to predict minimum values.

Based on a measured error rate, System A calculates and records a Q value every fifteen minutes. This results in 96 data points per day and 35,040 per year.

Assuming a Normal distribution, the $\overline{Q}-5\sigma$ value should represent a value, above which 99.999713%[1] of all values appear. This would indicate that over a one year period, following only the normal distribution, no naturally time varying values less than the $\overline{Q}-5\sigma$ calculation should be present [1].

$$.99999713 \times 35{,}040 = 35039.90 \qquad [1]$$

Data points below $\overline{Q}-5\sigma$ = 35,040 – 35,039.90 = 0.1

$$\approx 0$$

2 WHAT ARE TIME VARIATIONS?

There are many factors that are known to have an influence on the time varying nature of Q. These range from PMD through to environmental effects. It is supposed that if there are no other influences, the distribution of Q over a period of time will form a normal distribution.

Outliers are considered to be those points that exist outside of the normal distribution, being unusually large or small. These outliers need careful examination; they may be genuine results where the factors reach a highly unusual extreme or they could be the result of a non time-varying factor – such as a discrete fault. As such, outliers cannot be automatically discounted, but do need to be considered when analysing the data.

3 DATA

In preparing this paper, around 12 million data points were collected from System A, and a two year sample chosen from within. The data, from this 10G fully equipped system, with multiple terminal stations, was initially sanitised for obvious areas of spurious data. Straightforward examples include data obtained during a cable repair, where abnormal power feeding or a partial sampling window can lead to misleading Q values being recorded.

However, further examples can be harder to distinguish, and this leads to a certain degree of potential error. Most frequently, an individual hardware fault will sometimes lead to a drop in performance on one channel and a recorded Q value that is lower than would otherwise be present. The reliable identification of these values is not always straightforward. Operational records can be reconciled with the data but an accurate process involving such a quantity of data is prone to mistakes and omissions as well as significant resource requirements. For this reason, no further manipulation took place with the original data, with the understanding that suspect values might be included.

Where a suspect value is well within the normal distribution (close to the mean) there is very little effect on the overall distribution and the minimum Q value recorded is unaffected. However, when such an event causes a Q value that is lower than all other points in the distribution, simply looking at the lowest value can lead to misleading conclusions. To confirm validity of the stated minimum Q, a numerical summary of the frequency of all values present, starting from the minimum Q and increasing in 0.1dB increments over a 1dB range, was displayed. This easily allowed a quick confirmation of whether the minimum values were likely to be outliers (Figure 1) or the natural tail of the normal distribution (Figure 2).

Figure 1 shows a close up of one channel's distribution, and a few outliers can be seen at the lower end of the distribution. Without further analysis, it is not easy to state whether these are unlikely but feasible, naturally occurring values, or a collection of discrete failures.

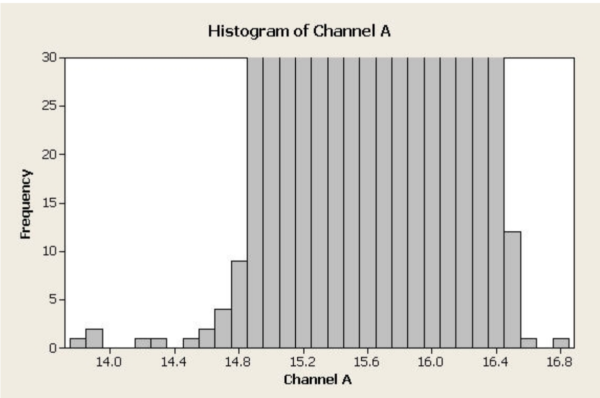

Figure 1: Close view of tail end, distribution A

Figure 2 shows something akin to a normal distribution where there are no outliers present after the natural tail of the distribution. In this case, it is more likely that the minimum value is a fair representation of the minimum value seen over the sample period.

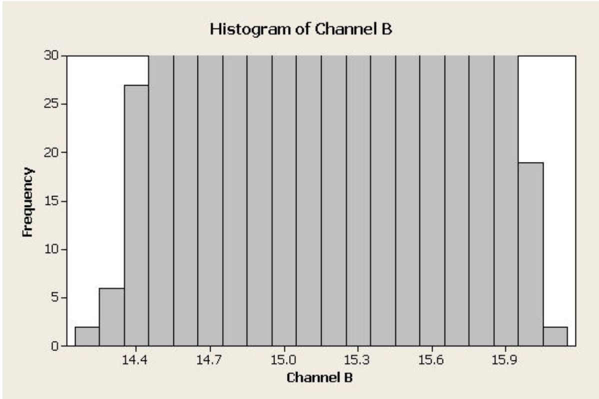

Figure 2: Close view of tail end, distribution B

For all channels like Channel A, the only way of being certain about the low values is to search the base data for the occurrence and then reference the date against Operation & Maintenance records to confirm if in fact it was due to a discrete failure. Even this method is not fool-proof as it is quite feasible for failures to manifest themselves over some time before it is flagged via operational systems as a fault.

4 NORMALITY

An important aspect of the analysis is whether the base data is Gaussian in distribution. In essence, it appears not when using standard industry tools, such as Minitab®[2], to perform a normality test. Even those distributions that appear "normal" to the eye, fail to meet the criteria that would support the hypothesis that the data comes from a normal distribution. The detailed discussion of normality testing is beyond the scope of this paper.

It can be noted though that a large subset of the channels are heavily skewed as there performance includes error free 15 minute periods. These naturally have an upper Q limit, beyond which no further measurements are possible.

5 METHOD

Using the data collected, a window of 1, 3, 5, 10, 21 and 30 days was used as theoretical confidence trials, with each sample period producing a calculated $\overline{Q} - 5\sigma$ value. These values were in turn compared to the minimum values seen during the following 12 months worth of real data. Figure 3 shows this procedure in more detail.

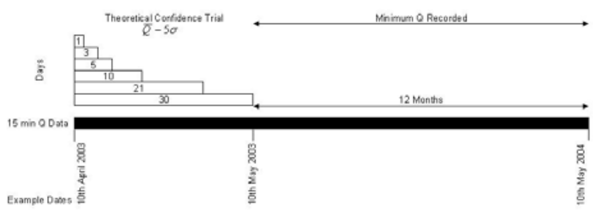

Figure 3: Schematic of data collection and analysis periods.

A year was chosen so as to be long enough that the data sample should cover all seasonal variations (should they be applicable) yet short enough to try and limit any effects from aging. As aging is considered as a separate line in power budgets, its effects are outside the time variations that the $\overline{Q} - 5\sigma$ calculation covers and so an assumption is made that the distribution does not reflect significant aging contributions.

The remaining data is considered to be a fair representation of the system operating under normal conditions with standard time varying factors and possible discrete faults.

6 HOW LONG SHOULD THE CONFIDENCE TRIAL BE?

As outlined before, the results of the various Confidence Trial lengths were analysed. Data from 3 different Terminal Stations was analysed by a rolling method for 8 consecutive periods. A single period is the one month maximum Confidence Trial window followed by a 12 month monitoring period. This period was then rolled one month forward and repeated to get variations in time on one channel. This method provided a total of 24 separate Confidence Trial periods, each measuring durations between one day and a month. These periods resulted in 1536 individual measurements when all channels are taken into account.

The results were graded into two categories. Firstly, where the Confidence Trial produced a $\overline{Q} - 5\sigma$ value that was lower than all data points in the following year, the result was considered Category 1. In this case the Trial predicted a value worse than was actually achieved. Category 2 was for results that were within 0.1dB of the minimum value, but were actually greater than it. The reason for doing this is due to the fact that the raw Q values are limited to 1 decimal place and so to allow for calculation accuracy and the rounding of the results, this error margin was applied.

The results show that, on average, the longer the Confidence Trial is, the more results there are that accurately predict the lowest value to be seen over a 12 month period. Table 1 summarises these results. It should be noted that in all cases, the population total was 1536. The percentages are derived from this maximum.

Duration (Days)	Category 1: No. of samples with minimum higher than predicted, i.e. prediction was worse case	Percentage of Cat.1 predictions (%)	Category 2: No. of samples with minimum within 0.1 dB lower than predicted	Total good (Cat.1 and Cat.2) predictions	Percentage of Total good predictions (out of 1536 samples) (%)
1	247	16.1	124	371	24.2
3	327	21.3	140	467	30.4
5	379	24.7	176	555	36.1
10	427	27.8	174	601	39.1
21	480	31.3	205	685	44.6
30	522	34.0	187	709	46.2

Table 1: Results of comparisons between Confidence Trial durations and samples from real data.

Allowing for the inclusion of Category 2 results, the accuracy even after a month is only around 50 %. Moreover, it can be seen that once the duration gets above 10 days, the incremental accuracy improvement for the additional time spent performing a Confidence Trial, is low.

A further check, is to confirm the $\overline{Q} - 5\sigma$ value produced by looking at one years worth of data and comparing this to the minimum value seen over the same period.

In this case, only 75% of the distributions satisfy this requirement, an indication of the non-normal distributions / outliers present.

In conclusion, it appears that the pay-off of the extended confidence trials, above 10 days, is limited if the sole purpose was to predict minimum Q.

7 AT WHAT POINT IS THE CONFIDENCE TRIAL MORE ACCURATE?

What has not been quantified so far, is the inaccuracy of the $\overline{Q} - 5\sigma$ values that are not represented within Category 1 or Category 2 definitions, over a 12 month sample period. To answer this, the difference between each of the Confidence Trial results and the minimum for the corresponding 12 month period was computed. By taking the x^{th} percentile, a value can be derived that indicates the limit that would be required to give x% accuracy. Values for 90%, 95% and 99% are shown in Table 2.

Duration (Days)	Margin Required (dB)		
	90% Confidence	95% Confidence	99% Confidence
1	0.97	1.19	1.65
3	0.86	1.07	1.47
5	0.80	1.03	1.39
10	0.74	0.90	1.31
21	0.66	0.83	1.22
30	0.64	0.80	1.22

Table 2: Margin required varying confidence limits.

These results show, based on this actual data set, the margin that would need to be subtracted from the Confidence Trial $\overline{Q} - 5\sigma$ value in order to give the % accuracy indicated above. For example, to ensure that 90% of the results for a 30 day Confidence Trial were less than or equal to the minimum value seen, the value required would be $\overline{Q} - 5\sigma$ - 0.64.

8 CONCLUSION

The use of $\overline{Q}-5\sigma$ as a tool during Confidence Trials is based on the distribution of Q being normal. Its ability to predict the worst case Q to be experienced over the medium term, with a long Confidence Trial, is roughly chance, although its ability to predict the minimum Q accurately is less.

However, it is noted that individual faults can affect the minimum Q seen over the medium term and without substantial investigation, it can be difficult to accurately rule data in or out.If the accuracy of $\overline{Q}-5\sigma$ is to be further understood in the future, the ability to correlate any abnormal event on any channel must be possible against the record of Q data so that the filtering of bad data can be performed more reliably.

It is also suggested that Suppliers might like to investigate ways of demonstrating that the $\overline{Q}-5\sigma$ figures will accurately represent the minimum to naturally occur over the medium term or propose alternative methods.

9 REFERENCES

[1] Notes on the Six Sigma Concept, William J Latzko, http://deming.ces.clemson.edu/pub/den/six_sig.pdf, 3 pages

[2] Minitab® is a registered trademark of Minitab Inc.