2007

The effect of data aggregation interval on voltage results

Sean Elphick
University of Wollongong, elpho@uow.edu.au

Victor Gosbell
University of Wollongong, vgosbell@uow.edu.au

Sarath Perera
University of Wollongong, sarath@uow.edu.au

The Effect of Data Aggregation Interval on Voltage Results

Sean Elphick, Vic Gosbell, Sarath Perera

Abstract-**For various technical and operational reasons, many power quality surveys are carried out using non-standard data aggregation intervals. The data aggregation interval is the time interval that rapidly sampled data is reduced to by the monitoring instrument for subsequent analysis and reporting. Some of the rationales for using non-standard data aggregation intervals include instrumentation limitations, memory restrictions, a belief that more insights may be obtained from data captured at faster aggregation intervals and dual use of instrumentation (such is the case for many smart revenue meters). There is much conjecture over the effect which the data aggregation interval will have on the final outcomes of a power quality survey. IEC61000-4-30 which is the international standard describing power quality monitoring methodology suggests 10 minute data aggregation intervals are appropriate for routine power quality monitoring of most power quality disturbances including magnitude of supply voltage. This paper investigates the variation observed for magnitude of supply voltage monitoring when data is captured at a range of data aggregation intervals.**

Index Terms—**Power Quality, Power Quality Indices, Data Aggregation**

I. INTRODUCTION

To report power quality it is necessary to reduce data sampled at high sampling rates down to a form which is useful without the loss of important detail. The method of reducing high speed data down to more useful data is known as aggregation and the time period over which the data is aggregated is called the data aggregation interval.

It is important to note the distinction between data aggregation interval and data sampling frequency. The sampling frequency is a basic function of the monitoring instrument and associated digital signal processing. Most modern instruments now sample at 256 samples per cycle or 12.8khz (or more) for continuous data thus exceeding the Nyquist requirements for sampling data up to the 50^{th} harmonic. The data aggregation interval is the time period over which the sampled data is combined to produce an average. For voltage measurement, most modern equipment measures the RMS value of the signal every half cycle. If the instrument is compliant with IEC 61000-4-30 [1] these half cycles values are then RMS averaged to a 10 cycle value which forms the basic building block for all aggregation to longer intervals. Common aggregation intervals include 3 seconds, 10 seconds, 10 minutes, 15 minutes, 1 hour and 2 hours; though in reality any aggregation interval is possible.

Once the most appropriate data aggregation interval has been determined, analysis and reporting of power quality data is generally performed by statistical analysis of data over specified time intervals. This time interval is often days or weeks. The data aggregation interval is very important because depending on the type of signal to be measured, too long an aggregation interval may result in the loss of important detail due to the RMS averaging processes. Too short an interval may result in copious amounts of data that is difficult to assess, may not be meaningful and presents a difficult storage problem if the data is to be retained. Therefore the aggregation interval much be chosen such that the amount of data to be analysed is reduced to manageable form whilst ensuring that sufficient detail is available to ensure a good indication of disturbance levels is achieved.

For most continuous disturbances, IEC61000-4-30, the international standard regarded as best practice for power quality monitoring, recommends 10 minute aggregation intervals for routine power quality monitoring surveys. There are many power quality surveys carried out using aggregation intervals other than the prescribed 10 minutes. Reasons for this vary but include: a perception that deeper insights will be obtained from results established using data sampled at a faster aggregation interval and dual use of instrumentation, for example smart revenue meters which generally aggregate data to 15 minute intervals.

Although standards define the recommended aggregation intervals for performing routine monitoring there is little indication in the standards, or other literature on the topic, concerning the impact that aggregating data at non-standard intervals will have on the results of routine monitoring. In [2] it is demonstrated that using different aggregation intervals has the potential to mask otherwise important voltage behaviour, though the paper does not give any specific recommendations as to the most appropriate aggregation interval to use. The method suggested in clause A.6.2.2 of IEC61000-4-30 for conducting routine magnitude of supply voltage surveys is assessment of 95^{th} percentile values of 10 minute voltage data over one week. In spite of the fact that the standard calls for the 95^{th} percentile statistical confidence level to be used for analysis of data, other statistical levels are often discussed and may be useful in some cases. These include statistics such as the maximum (100^{th} percentile) and the 99^{th} percentile.

Sean Elphick is with the Integral Energy Power Quality and Reliability Centre, School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, Wollongong, NSW, Australia 2522. Email: sean@elec.uow.edu.au
Vic Gosbell is with the Integral Energy Power Quality and Reliability Centre, School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, Wollongong, NSW, Australia 2522. Email: vgosbell@uow.edu.au
Sarath Perera is with the Integral Energy Power Quality and Reliability Centre, School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, Wollongong, NSW, Australia 2522. Email: sarath@uow.edu.au

In the case of voltage, analysis is complicated by the fact that the optimum value for voltage is not zero but the nominal voltage which is contained within a double sided band. Therefore, to fully quantify voltage performance two statistics are necessary, one for the high end of the range and one for the low end. Thus voltage may be described by a maximum and a minimum (0 percentile) or other statistics such as the 95th percentile and the 5th percentile.

This study quantifies the effect that using different aggregation intervals will have on the statistical results over a one week period. This allows conclusions to be made regarding how aggregation interval influences the reported voltage magnitude values. It should be noted that this study only addresses this question for routine monitoring purposes. For troubleshooting or fault investigations, other aggregation intervals may be more appropriate. Three sets of statistics to characterise voltage are concentrated on, namely, the maximum and minimum, the 99th percentile and 1st percentile, and the 95th percentile and 5th percentile. The aggregation intervals to be tested are 30 seconds (which was the shortest interval over which a meaningful amount of data could be compiled), 1 minute, 10 minutes (recommended interval in IEC61000-4-30), 15 minutes (corresponding to the basic revenue metering interval which is used in many smart tariff meters) and 1 hour.

II. TEST DATA

The test data used in this study has been collected by The University of Wollongong during various power quality projects. All data used was recorded by monitoring instrumentation employing data aggregation intervals of 30 seconds or less.

There are 9 distinct sites which have provided data for this study. These sites are a mixture of low voltage and medium voltage sites. Of these sites some provided data for one week whilst others provided data for multiple weeks. For sites with data spanning multiple weeks, data was chosen during different times of the year in order to attempt to quantify the seasonal effects on the results. For the purposes of this study, where data was collected over multiple contiguous weeks for one site, the data corresponding to each week is treated as if it is an independent site. For these sites the naming convention adopted in this paper is for the site to have one numerical identifier and the weeks to be further numbered. For example Site 1 which has three weeks of data, will be named Site 1 Week 1, Site 1 Week 2 and Site 1 Week 3. Where necessary for ease of graphing, week has been reduced to W resulting in Site 1 W1 for Site 1 Week 1 and so on.

All of the sites used in this study are strong sites, meaning that they are located close to or at a transformer. Details of the sites supplying data to this study are outlined below:-

- Site 1: Medium voltage. 3 weeks of data.
- Site 2: Low voltage. 3 weeks of data.
- Site 3: Medium voltage. 3 weeks of data.
- Site 4: Medium voltage. 4 weeks of data.
- Site 5: Medium voltage. 3 weeks of data.
- Site 6: Medium voltage. 3 weeks of data.
- Site 7: Zone Substation which supplies 3 low voltage sub-sites, Site 7a, 7b and 7c. All sites have 1 week of data.
- Site 8: Zone Substation which supplies 4 low voltage sub-sites, Site 8a, 8b, 8c and 8d. All sites have 1 week of data.
- Site 9: Medium voltage. 1 week of data.

III. TESTING PROCEDURE

The process of aggregating data from short time intervals to longer time intervals using an RMS averaging process results in smoothing of the data and the loss of high frequency components as the data aggregation interval gets longer and longer.

Before examining the test procedures used to assess the variation of data across aggregation intervals it is worth reviewing methods of statistical analysis of data and statistical confidence levels. A statistical confidence level describes a value for which the data will less than or equal to for a certain percentage of the time.

As an example consider the 95th percentile. This is the value for which the data will be less than or equal to for 95% of the time. Of course the timeframe over which the statistical confidence level is determined is important and there has been some debate on this topic, however, it is beyond the scope of this study to discuss these ideas. If the 95th percentile level over 1 week is considered it can be calculated that the 95th percentile level will exclude 8.4 hours worth of data from the week. This equates to the loss of 9 hourly intervals, 34 fifteen minute intervals, 51 ten minute intervals, 504 one minute intervals and 1008 thirty second intervals.

Fig 3.1 shows the one week trend of data from Site 2 Week 3 for some of the data aggregation intervals under study in this paper. It can be seen that as data is aggregated to longer and longer intervals there is a noticeable smoothing effect. That is, although the basic shape of the trend is preserved, there is a loss of high frequency peaks and troughs. This is particularly apparent when the data is aggregated from 1 minute intervals up to 10 minute intervals. It can be seen that the spikes seen in the 30 second and 1 minute data aggregation trends are not carried through to the 10 minute trend. This smoothing of data results in short term values which will be higher (at the top end of the voltage range) and lower (at the bottom end of the voltage range) than data aggregated at longer intervals. The exact variation between data aggregated to the different intervals is discussed below.

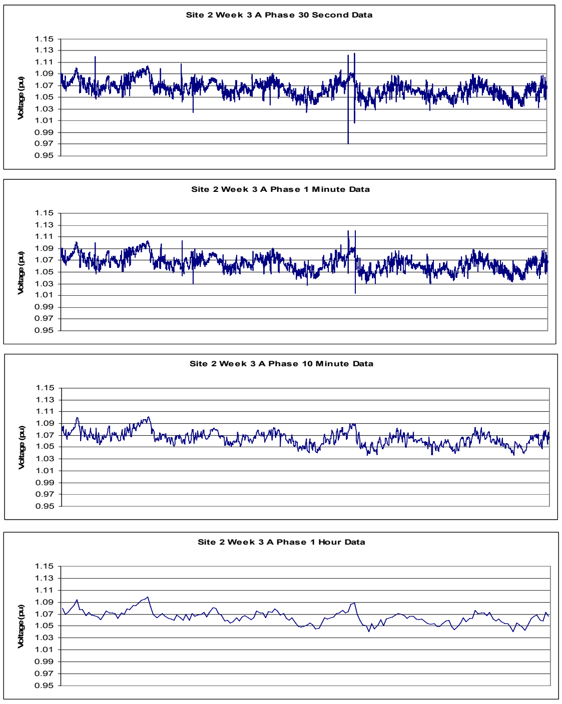

Fig 3.1: One Week Voltage Trend for Site 5; Data Aggregated to 30 Second, 1 Minute, 10 Minute and 1 Hour Intervals

Three basic statistical confidence levels are examined in this paper. These are (a) the maximum, (b) the 99th percentile and (c) the 95th percentile. As the voltage data is subject to an upper and a lower limit, it is necessary to calculate statistical confidence levels to assess both the upper and the lower end of the voltage scale. Thus the lower end of the voltage range is defined by statistical confidence levels which are symmetrical to the three given above. These are (a) the minimum, (b) the 5th percentile and (c) the 1st percentile.

Once a uniform data set was realised the testing procedure was relatively straight forward. Using the 30 second data as a base, the data was further aggregated to produce values for each of the aggregation intervals under test. Once this was achieved, the first stage of assessing the variations in values calculated using different aggregation intervals involved calculating statistical levels for each of the test aggregation intervals at each site. That is calculation of the maximum, 99th percentile, 95th percentile, 5th percentile, 1st percentile and minimum values over a one week period for each site using data aggregated to: -

- 30 seconds
- 1 minute
- 10 minutes
- 15 minutes
- 1 hour

For example for the case of the maximum there will be 5 values for each site, that is the maximum 30 second value, 1 minute value, 10 minute value, 15 minute value and 1 hour value. As an example of the variation seen when statistical confidence levels are calculated for one week using different aggregation intervals, Appendix A shows a graph of the maximum values obtained for sites 1, 2, 3 and 4 for each test aggregation interval.

Once these statistics have been calculated for each site the variation of each statistical level across the 5 different aggregation intervals under study can be determined. These variations may then be used as the basis for calculating the characteristic variations across the 5 different aggregation intervals for all sites.

IV. RESULTS

A. Variation between 30 Second and 1 Hour Data Aggregation Intervals

Once data has been statistically analysed it is possible to take measures of the variation of the statistical measures for each site. For example, each site will have a maximum value for each test aggregation interval. It is then possible to calculate the variation of the maximum values across the aggregation intervals. If this process is repeated for each site there will be 27 (one for each site) variations calculated. From these 27 variations, further statistics can be calculated such as the maximum of the variation of maximum values, i.e. the maximum of the 27 variations obtained for the maximum at each site, and the average value of the variation of maximums.

Fig 4.1 shows the variation between the shortest aggregation interval, 30 seconds, and the longest aggregation interval, 1 hour for each of the tested statistical measures, calculated using the method described above. Using the shortest and longest intervals give the absolute maximum variation that will be seen and represents a worst case scenario.
It can be seen that most variation of the test aggregation intervals occurs for variations in minimum levels, followed by variations in maximum levels. Fig 4.1 also indicates that there is very little variation in the test aggregation intervals for the statistical confidence levels; 99th percentile, 95th percentile, 5th percentile and 1st percentile. Detailed analysis of the variations across the test data aggregation intervals is outlined below.

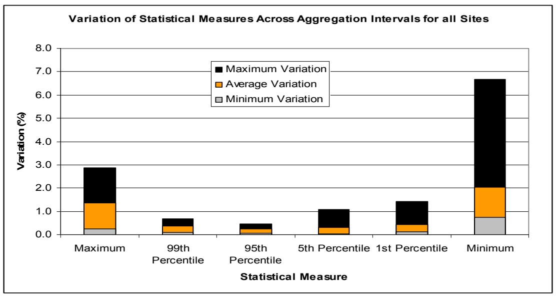

Fig 4.1: Variation of Statistics across Test Aggregation Intervals

A1. Variation of Maximum and Minimum Values

Fig 4.1 shows that there is considerable variation of maximum values across the different aggregation intervals. The variation between the maximum maximum value at a site which occurs for 30 second data and the minimum maximum value which occurs for 1 hour data for all sites was 2.8%. This means that if data was assessed using 30 second data aggregation intervals the value reported would be 2.8% higher than the value reported if the data was assessed using data aggregated to 1 hour intervals. Given that the nominal voltage range is 12% for low voltage and 10% for medium voltage, this

2.8% increase represents 23% and 28% of the voltage ranges respectively which is a significant figure. High variation in the maximum values is an expected result due to the fact that the maximum is a quite volatile statistic and it is for this reason that maximum values are not generally used for assessment of site performance. With respect to aggregation intervals it is unsurprising that the 30 second maximum is considerably higher than the 1 hour maximum as any rapid changes occurring on the 30 second time-scale would need to persist for quite some time to have any impact on the 1 hour value. This indicates that maximum values are occurring randomly and rarely persist long enough to have an impact on the longer term aggregation intervals.

The average variation between the maximum values across aggregation intervals was found to be 1.4%. This represents a 50% decrease on the maximum variation of the maximum values. The large difference between the average variation of maximum values and the maximum variation of maximum values indicates either that there are a few sites which have large maximum variations and some sites which have very small maximum variations or that there is a constant distribution of maximum variations across all sites with some sites being large, some average and some small. Analysis of the data as shown in Fig 4.2 which shows the distribution of the maximum variations proves the second case to be true, that is there is a constant distribution of maximum variations. This again attests to the random nature of maximum values.

For variation of the minimum values, similar results are observed as for the variation of maximum values although the variation in minimum values is considerably higher than that seen for maximum values. The same reasoning regarding the random nature of minimum values as was applied for maximum values can be used to explain the high variation of the minimum values.

The maximum variation of minimum values across all sites was 7% of the nominal voltage which is very large. This value represents more than 50% of the nominal voltage range for both low voltage and medium voltage. This suggests that the data aggregation interval will play a large part in the results of surveys if the minimum value is used as an assessment criterion.

The average variation of minimum values was found to be 2% and the minimum 0.7% both of which are considerably larger than the corresponding maximum values. Fig 4.2 which shows the variation of the maximum values and the minimum values across the test aggregation intervals for each site clearly shows that there is more variation in minimum values than maximum values.

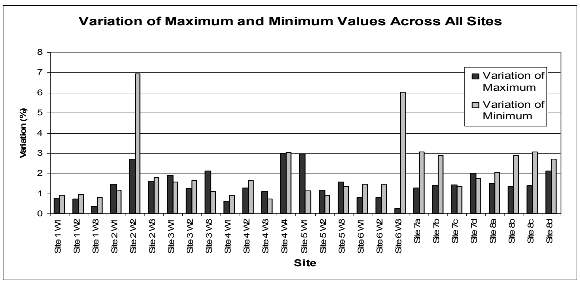

Fig 4.2: Variation of Maximum and Minimum Values

A2. Variation of Other Statistical Measures

Fig 4.1 clearly shows that there is significantly less variation seen across data aggregation intervals for the statistical confidence levels examined. This is due to the fact that application of statistical confidence levels will exclude a number of the most extreme values. The fact that the values obtained when statistical confidence levels are applied vary little regardless of data aggregation interval confirms the randomness of maximum values as discussed above. Voltage readings at most sites will be characterised by many values clumped close together along with some very rarely occurring outlying values. These outliers may be due to unusual events on the network, are not persistent, and are often unrepeatable. This is the reason that most standards avoid using the maximum value for comparison with limits or planning levels. Taking a statistical confidence level such as the 99th percentile eliminates the most extreme of these outliers and gives a value which is more likely to be repeatable.

For the 99th percentile the maximum variation between 99th percentile values for the test data aggregation intervals was found to be 0.68%. This is 75% less than the corresponding value seen for maximum values. In addition this value represents only 5.6% of the low voltage nominal range and 6.8% of the medium voltage nominal range and is small enough to conclude that the aggregation interval used to calculate 99th percentile values over a week will have little impact on the outcome of the statistical analysis.

The average variation between 99th percentile values across all sites was found to be 0.37% and the minimum variation was found to be 0.11%.

Similar conclusions as were made for the 99th percentile values can be made for the 95th percentile values. There is even less variation in 95th percentile values than there was for 99th percentile values. Once again this small variation indicates that the 95th percentile level will be similar regardless of the base aggregation interval that is used for calculation of the statistic. In fact the maximum variation in 95th percentile values across the test data aggregation intervals for all sites was found to be only 0.5% while the average variation was found to be 0.27% and the minimum 0.06%. This indicates that if a 95th percentile value is to be used as the reporting statistic there is no need to aggregate data more frequently than the 10 minute interval prescribed in IEC61000-4-30.

The 5th percentile results are similar to the 95th percentile and this is expected due to the symmetry of the statistics. The maximum variation in 5th percentile readings across the test aggregation intervals was found to be 1.4% which is larger than the corresponding value for the 95th percentile. However, the average variation in 5th percentile values across all sites was found to be 0.3% and the minimum variation in 5th percentile values was found to be 0.04% voltage both of which are smaller than the corresponding 95th percentile values.

1st percentile values across all sites are somewhat higher than the variation seen for 99th percentile values. This follows the trend seen for the 5th percentile and minimum which indicates that the statistics which describe the lower end of the voltage scale (5th percentile, 1st percentile, minimum) have larger variation than the statistics which described the upper end of the voltage scale (maximum, 99th percentile, 95th percentile). This indicates that there are more rapid changes and/or changes of larger magnitude in voltage occurring at the low end of the voltage scale as opposed to the high end. The maximum variation in 1st percentile values across the test data aggregation intervals was found to be 1.5%. This value is 233% larger than the corresponding 99th percentile value, but is probably still small enough not to justify the use of shorter aggregation intervals which will result in much more data to be analysed and stored. The average variation in 5th percentile values across all sites was found to be 0.45% and the minimum variation was found to be 0.13% both of which are comparable to the corresponding 99th percentile values.

B. Variation between 30 Second and 10 Minute Values

The 10 minute data aggregation interval is referred to in many standards as the aggregation interval which should be used for routine power quality monitoring. This section describes the variations seen when data aggregated to 10 minute intervals is reported as opposed to data aggregated to 30 second intervals. Fig 4.3 shows the variation between 30 second and 10 minute aggregation intervals.

Fig 4.3 indicates that there is significant variation between 30 second data and 10 minute data for the maximums and the minimums. Once again it appears that there are a few sites which are outliers, characterised by the maximum variation being significantly larger than the average variation. This reflects the trend seen in the above sections for 30 second data and 1 hour data. It can be seen that there is very little variation between the data aggregated over the two different intervals if statistical confidence levels as opposed to the volatile maximum and minimum statistics are used.

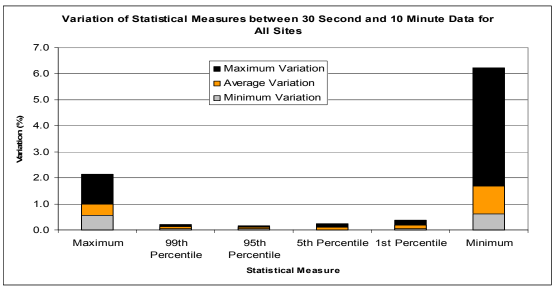

Fig 4.3: Variation between 30 Second and 10 Minute Data for Test Statistics

C. Variation between 10 minute and 15 minute values

Many utilities use smart revenue meters for collection of power quality data. These devices are useful in that they are often installed at important locations to collect revenue data and the addition of some basic power quality functionality is often a fiscally attractive method of obtaining power quality data. One of the drawbacks of these instruments is that many either are not configured for measuring at 10 minute aggregation intervals or are not able to monitor power quality at the 10 minute interval recommended in IEC61000-4-30. Most smart revenue meters aggregate data at 15 minute intervals which is the standard revenue metering period.

Analysis has been performed to quantify the difference between aggregating data at 10 minute intervals and 15 minute intervals. The results are quite conclusive. For the most volatile statistical indices, namely the maximum and the minimum, the maximum variations between a 10 minute value and a 15 minute value seen at any site were found to be 0.36% and 0.63% respectively. The average variation between a 10 minute value and a 15 minute value for maximum values was found to be 0.12% and 0.13% for minimum values. These variations are very small and indicate that, for voltage at least, data aggregated at 15 minute intervals will be so close to the value of data aggregated at 10 minute intervals as to be almost identical.

V. DISCUSSION

A. Recommended Aggregation Interval for Voltage Reporting

In almost all cases voltage is reported using a statistical confidence level. This study shows that there is little variation between values calculated for statistical confidence levels regardless of the data aggregation interval. This indicates that there is little gain in insights obtained when data is aggregated at short intervals. This suggests that there is no convincing reason to aggregate voltage data any faster than the 10 minute interval specified by IEC61000-4-30. This is convenient due to the fact that routine power quality monitoring can produce enormous amounts of data particularly where harmonics are monitored and it is important to try and keep data amounts as low as possible both to simplify analysis and ease data storage burdens. In addition it is not clear what potential gain utilities in particular will achieve by routine voltage supply magnitude monitoring at data aggregation intervals faster than intervals in the order of minutes as it is only changes in voltage supply magnitude of these orders that the utility can control through network operations anyway.

B. Applying Limits to Maximum and Minimum Values

The high variation in maximum and minimum values noted in this study is consistent with these values being an inconsistent measure and is the reason why they are not often recommended for comparison with limits or planning levels. Although volatile, it may not be possible to ignore maximum and minimum values entirely when voltage supply magnitude is measured. A limit may need to be placed on the absolute levels on which these values can reach in order to prevent damage to equipment. This limit may need to be specified as part of a voltage standard or may be limited by sag and swell thresholds (which are generally defined as ±10% of the nominal voltage). In spite of the volatility of these statistical measures and the fact that the variation across aggregation intervals is significant it should be noted that the variation at any site, while large, is not extreme. If a limit is to be imposed on these measures it must have a range suitably wide to take into account the values may only occur very rarely and may be due to abnormal operating conditions.

VI. CONCLUSIONS

Data has been analysed from 9 distinct sites giving a total of 27 weeks worth of data aggregated at 30 second intervals. This data has been used to determine the effect of data aggregation interval on reported voltage magnitude levels.

The study shows that there is little variation between the weekly values for 99^{th}, 95^{th}, 5^{th} and 1^{st} percentile values at each site regardless of the aggregation interval used in the calculation of the confidence intervals. This result indicates that little additional insight will be achieved by aggregating data at intervals faster then the 10 minute interval specified by IEC61000-4-30.

Analysis of maximum and minimum data shows large variation across aggregation intervals with the values for shorter aggregation intervals. This indicates that there are rapid changes in voltage occurring over very short time periods which are not persistent enough to influence longer aggregation intervals. Thus, if maximum and minimum values are to be used as assessment criteria for supply voltage magnitude, the aggregation interval will play a significant role in the outcome of the assessment and needs to be carefully specified.

Many utilities use smart tariff meters aggregating data at 15 minute intervals for power quality data collection. Analysis of voltage data aggregated at 10 minute intervals and data aggregated at 15 minute intervals showed the discrepancies between the statistical parameters calculated in this study to be insignificant.

VII. REFERENCES

[1] *IEC Standard for Electromagnetic Compatibility (EMC) – part 4-30: Testing and Measurement Techniques – Power Quality Measurement Methods,* IEC61000-4-30, 2000.

[2] Daniele Gallo, Roberto Langella and Alfredo Testa, *"Predicting Voltage Stress Effects on MV/LV Components"*, Power Tech Conference Proceedings, 2003 IEEE Bologna, Volume 2, 23-26 June 2003.

VIII. BIOGRAPHIES

Sean Elphick graduated from the University of Wollongong with a BE (Elec) degree in 2002. He commenced employment with the Integral Energy Power Quality Centre in 2003. Initially employed to work on a Strategic Partnerships with Industry - Research and Training Scheme (SPIRT) project dealing with power quality monitoring and reporting techniques. His current activities include delivery of the Long Term National Power Quality Survey, a first of its type in Australia as well as various other power quality related research and consulting projects.

Vic Gosbell obtained his BSc, BE and PhD degrees from the University of Sydney. He has held academic positions at the University of Sydney and the University of Wollongong where he became the foundation Professor of Power Engineering. He is now an Honorary Professorial Fellow and Technical Advisor to the Integral Energy Power Quality and Reliability Centre. He is currently working on harmonic management, power quality monitoring and standards. He is a member of Australian standards and CIGRE sub-committees and is a Fellow of the Institution of Engineers, Australia.

Sarath Perera graduated from the University of Moratuwa with a BSc (Eng) specializing in Electrical Power Engineering. He obtained his MEngSc from the University of NSW and PhD from the University of Wollongong. He has been formerly attached to University of Moratuwa as a Lecturer and is now an Associate Professor at the University of Wollongong and the Technical Director of the Integral Energy Power Quality and Reliability Centre.

APPENDIX A. MAXIMUM VOLTAGE LEVELS FOR EACH TEST AGGREGATION INTERVAL FOR SELECTED SITES

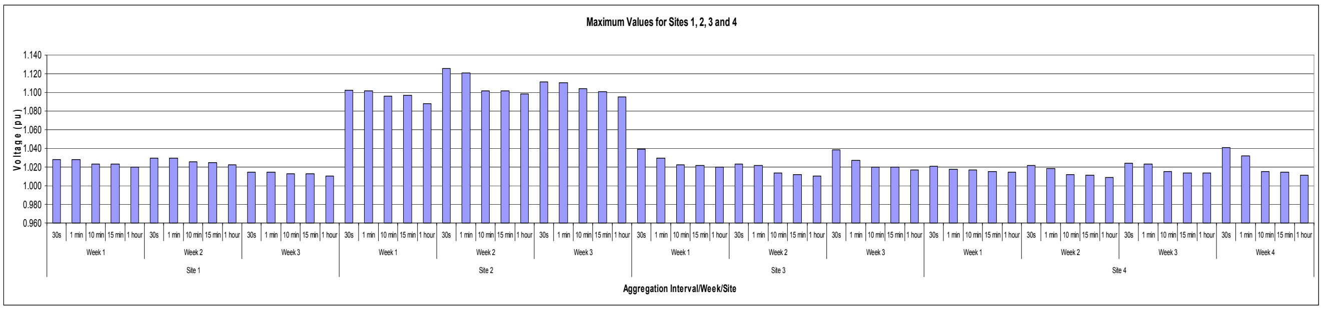

Fig A1: Maximum Values across Test Aggregation Intervals for Selected Site