1-1-2006

Analyzing harmonic monitoring data using data mining

Ali Asheibi
University of Wollongong, ali_asheibi@uow.edu.au

David A. Stirling
University of Wollongong, stirling@uow.edu.au

Danny Sutanto
University of Wollongong, soetanto@uow.edu.au

Recommended Citation
Asheibi, Ali; Stirling, David A.; and Sutanto, Danny: Analyzing harmonic monitoring data using data mining 2006, 63-68.
https://ro.uow.edu.au/infopapers/1802

Analyzing harmonic monitoring data using data mining

Abstract

Harmonic monitoring has become an important tool for harmonic management in distribution systems. A comprehensive harmonic monitoring program has been designed and implemented on a typical electrical MV distribution system in Australia. The monitoring program involved measurements of the three-phase harmonic currents and voltages from the residential, commercial and industrial load sectors. Data over a three year period has been downloaded and available for analysis. The large amount of acquired data makes it difficult to identify operational events that impact significantly on the harmonics generated on the system. More sophisticated analysis methods are required to automatically determine which part of the measurement data are of importance. Based on this information, a closer inspection of smaller data sets can then be carried out to determine the reasons for its detection. In this paper we classify the measurement data using data mining based on clustering techniques which can provide the engineers with a rapid, visually oriented method of evaluating the underlying operational information contained within the clusters. The paper shows how clustering can be used to identify interesting patterns of harmonic measurement data and how these relate to their associated operational issues.

Keywords

Analyzing, harmonic, monitoring, data, using, data, mining

Disciplines

Physical Sciences and Mathematics

Publication Details

A. Asheibi, D. A. Stirling & D. Soetanto, "Analyzing harmonic monitoring data using data mining," in Australian Data Mining Conference, 2006, pp. 63-68.

Analyzing Harmonic Monitoring Data using Data Mining

Ali Asheibi, David Stirling, Danny Soetanto

Integral Energy Power Quality and Reliability Centre
School of Electrical, Computer and Telecommunications Engineering
University of Wollongong
Northfields Ave, Wollongong 2522, NSW

Email: atma64@uow.edu.au

Abstract

Harmonic monitoring has become an important tool for harmonic management in distribution systems. A comprehensive harmonic monitoring program has been designed and implemented on a typical electrical MV distribution system in Australia. The monitoring program involved measurements of the three-phase harmonic currents and voltages from the residential, commercial and industrial load sectors. Data over a three year period has been downloaded and available for analysis. The large amount of acquired data makes it difficult to identify operational events that impact significantly on the harmonics generated on the system. More sophisticated analysis methods are required to automatically determine which part of the measurement data are of importance. Based on this information, a closer inspection of smaller data sets can then be carried out to determine the reasons for its detection. In this paper we classify the measurement data using data mining based on clustering techniques which can provide the engineers with a rapid, visually oriented method of evaluating the underlying operational information contained within the clusters. The paper shows how clustering can be used to identify interesting patterns of harmonic measurement data and how these relate to their associated operational issues.

Keywords: Harmonics, Power Quality, Monitoring system, Data Mining, Classification, Clustering, Segmentation.

1 Introduction

With the increased use of power electronics in residential, commercial and industrial distribution systems, combined with the proliferations of highly sensitive micro-processor controlled equipment, more and more distribution customers are sensitive to excessive harmonics in the supply system (Heydt 1998), some even leading to failure of equipment. An increasing number of electric distribution network service providers are installing harmonic monitoring equipment to measure the three Phase harmonic voltage and current waveforms in their power system to detect and mitigate the harmonic distortion problems (Shuter, Vollkommer et al. 1989; Duggan and Morrison 1993; Lachaume, Deflandre et al. 1993; Morrison and Duggan 1993; Khan 2001; McGranagahn 2001).

Recently, a harmonic monitoring program was designed and implemented in a medium voltage (MV) distribution system in Australia (Gosbell, Mannix et al. 2001; Robinson 2003). The monitoring involved simultaneous measurements of the three-phase harmonic current and voltage from the residential, commercial, and industrial load sectors. The simultaneous measurements of three-phase harmonic currents and voltages from the different load sectors allow for the effect on the net distribution system harmonic voltage and current to be determined. The coordinated approach in obtaining the results has overcome some of the problems with synchronizing and reporting data (Emanuel and et al 1993; Sabin, Brooks et al. 1999).

An enormous amount of data over a three year period has been downloaded and available for analysis. However, it is difficult to analyze the data using visual inspection of the acquired voltage and current waveforms. It is also difficult to identify operational issues that generate the harmonics produced at varying operation time. A more sophisticated analysis method is required to automatically segment the data into manageable data set for analysis to understand the causes and effects of the harmonics obtained. These can then be used for resolving possible existing problems and to predict future problems. This implies that some decisions regarding data selection and acquisition must be made.

In this paper, a data mining tool is used for the automatic segmentation of the harmonic database. Segmentation (or clustering) is the discovery of similar groups of multidimensional records in databases and is a powerful mining tool. The data mining tool is based on the successful AutoClass (Cheeseman and Stutz 1995) and Snob research programs (Oliver, Baxter et al. 1996) and uses mixture models (McLachlan 1992) to represent clusters. The tool allows for the automated selection of the number of clusters and for the calculation of means, variances and relative abundance of the clusters. The paper describes how the data mining tool is used to search and analyze the multidimensional patterns on a cluster basis for the three-phase harmonic voltage and current data acquired from the harmonic monitoring program. By observing the data in each cluster, several very interesting operational data can be deduced.

The paper will first define harmonics, describe the design and implementation of the harmonic monitoring program and the data obtained. Results from the harmonic monitoring program are then analyzed using the data mining tool. The paper discusses the significance of the clusters obtained and how the associated operational conditions can be deduced from the resultant clusters. Several observations on the use of such data mining tools for analyzing large amounts of harmonic data are also discussed.

2 Understanding Harmonics

Ideally, the waveforms of the voltage supplied by the utility and the current utilized by consumers are perfect 50HZ sine wave. However, in practice these waveforms are deformed, producing multiple frequencies other than the 50Hz sine wave This phenomenon of wave deformation due to multiple frequencies is often referred to as '*harmonic distortion*'. The sources of the harmonic are electronic based equipment, such as computers, fax machines, TVs etc and arcing devices, such as arc furnaces, arc welders and dischargeable lighting. Other sources of harmonic are saturable devices like transformers and motors (Shwehdi, Mantawy et al. 2002). Harmonic distortion can cause serious long term and short term problems to power systems, as well as to consumers. Harmonics can cause capacitor failure, overheating of neutral conductors and false tripping of electrical equipment in the utility. Harmonic losses in industrial systems can increase the operational cost and decrease the useful life of the system equipment (Carpinelli, Caramia et al. 1996). There are several practical mitigation methods, such as introducing filters, modifying loads and adjusting the frequency response of the system. For harmonic management of a network, it is mandatory to carry out careful harmonic monitoring analysis in order to determine the harmonic levels at any point in the power systems or at customer sites. The next section explains the steps of the harmonic monitoring program that has been carried out in this study.

3 Harmonic Monitoring Program

A harmonic monitoring program (Gosbell, Mannix et al. 2001; Robinson 2003) utilized a typical MV distribution system in Australia in a typical 33/11kV zone substation that supplies ten 11kV radial feeders. The zone substation is supplied at 33kV from the bulk supply point of a transmission network. Figure 1 gives the layout of the zone substation and feeder system for the harmonic monitoring program.

Seven monitors were installed, a monitor at each of the residential, commercial and industrial sites (site ID 5-7), a monitor at the sending end of the three individual feeders (site ID 2-4) and a monitor at the zone substation incoming supply (site ID 1). Sites 1-4 in Figure 1 are all within the substation at the sending end of the feeders identified as being of a predominant load type. Site 5 was along the feeder route approximately 2km from the zone substation, feeds residential area. Site 6 supplies a shopping centre with a number of large supermarkets and many small shops. Site 7 supplies a factory manufacturing paper product such as paper towels, toilet paper and tissues. Based on the distribution customer details, it was found that site 2 comprises 85% residential and 15% commercial, site 3 comprises 90% commercial and 10% residential and site 4 comprises 75% industrial, 20% commercial and 5% residential. The monitoring equipment used is the EDMI Mk3 Energy Meter from Electronic Design and Manufacturing Pty. Ltd. as shown in Figure 2 (EDMI 2000). All three line-to-neutral voltages and line currents were recorded at the LV locations. The memory capabilities of the above meter at the time of purchase limited recordings to the fundamental current and voltage in each phase, the current and voltage THD in each phase, and three other individual harmonics in each phase.

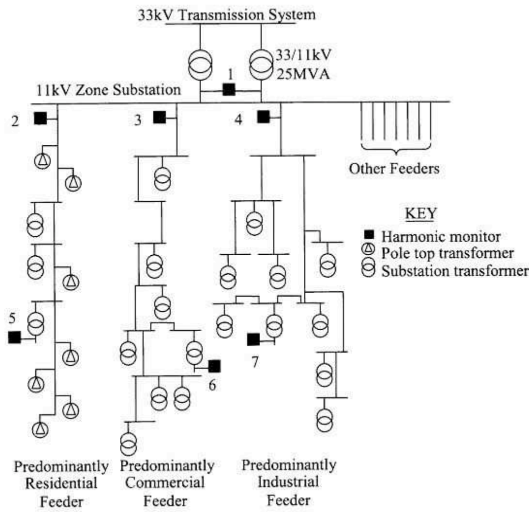

Figure 1: Single line diagram illustrating the zone distribution system

For the harmonic monitoring program, the harmonics chosen to be recorded were the 3rd, 5th and 7th harmonic currents and voltages at each monitoring site, since these are the most significant harmonics. The memory restrictions of the monitoring equipment dictated that each parameter was recorded only every 10 minutes. The data retrieved from the harmonic monitoring program spans from August 1999 to December 2002.

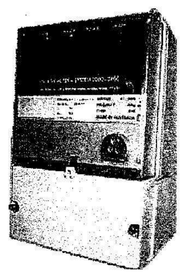

Figure 2: EDM1 2000-04XX Energy Meter

Figures 3-4, show a typical output data of the fundamental, 3rd, 5th and 7th harmonic currents in Phase 'a' at sites 1 and 2, taken from 6 -19 January 2001. It is obvious that for the engineers to realistically interpret such large amounts of data, it will be necessary to cluster the data into meaningful segments.

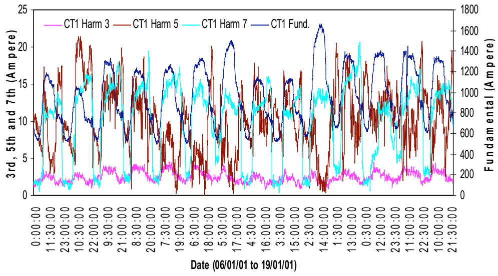

Figure 3: Phase 'a' Fundamental, 3rd, 5th and 7th Harmonic current waveforms in site 1 (zone substation)

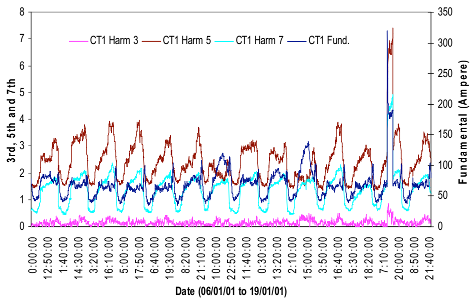

Figure 4: Phase 'a' Fundamental, 3rd, 5th and 7th Harmonic current waveforms in site 2 (residential)

4 Using Unsupervised Clustering with MML

Unsupervised clustering is based on the premise that there are several underlying classes that are hidden or embedded within a data set. The objective of such processes is to identify an optimal model representation of these intrinsic classes, by separating the data into multiple clusters or subgroups.

The partitioning of data into candidate subgroups is usually subject to some objective function like a probabilistic model distribution, e.g. Gaussian. From any arbitrary set of data, several possible models or segmentations might exist with a plausible range of clusters. Accordingly an appropriate evaluation scheme, such as Minimum Message Length (MML), or, Minimum Description Length (MDL) encoding, is used to evaluate each successive set of segmentations and monitor their progression towards a globally best model. This methodology is also known as finite mixture modelling or intrinsic classification (Wallace 1968; Wallace and Dowe 1994; R. Xu and D. Wunsch II 2005) . Algorithms such as "SNOB" or "Auto Class", which employ an MML approach have been found to form clusters that are statistically distinct (unique) within the multivariate parameter space of the data. The specific software used in this work was ACPro, one of several data mining algorithms that have yielded valuable insights for a range of scientific, and industrial real-time data (Oliver, Roush et al. 1998; Stirling and Zulli 2004). Finite mixture models provide a more formal (probabilistic-based) mechanism with which to fit arbitrary complex probability density functions (pdf's) of the data. In addition, from a practical perspective, such models also provide relief from the inherent constraints (priors and initial conditions) that accompany heuristic (distance) methods such as k-means or hierarchical agglomerative approaches (M. A.T. Figueiredo and A. K. Jain 2002).

5 Results and outcomes

A specialized data mining software package for the automatic segmentation of databases, ACPro, was used in this work. ACPro is essentially an unsupervised clustering that utilize minimum message length (MML) (Oliver, Baxter et al. 1996) encoding metric. It has the ability to discover similar groups of records in the database in the form of clusters. ACPro was applied to the measured harmonic data from the monitoring program for the test system in Figure 1. The data from different sites (sites 1, 2, 3 and 4) were used as input data to the software and 5 different clusters (s0, s1, s2, s3 and s4) each with specific abundance, mean and standard deviation were obtained. The clusters were then sorted in ascending order based on the mean value of the fundamental current, such that cluster (s0) is associated with the off peak load period and cluster (s4) is the cluster related to the on-peak load period. Figure 5(a) shows these clusters superimposed on the fundamental current waveform at site 3 (commercial). Figure 5(b) shows the abundance, mean and standard deviation for the clusters of the three harmonic currents per one phase.

By observing how the measured data are classified into various clusters, the Utility Engineer can more readily deduce the power quality event that may have trigger a change from one cluster to another. To confirm that, other available data in the utility can be used, such as temperature and reactive power measurements.

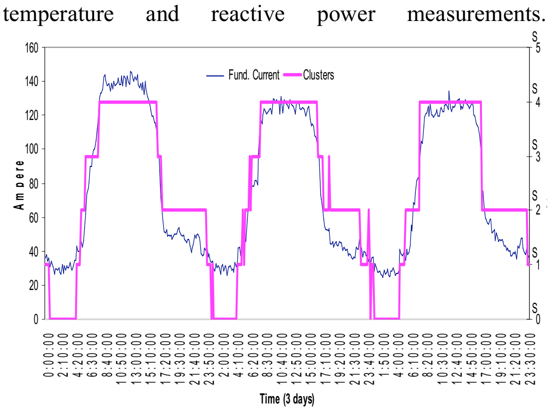

Figure 5(a): The clusters obtained superimposed on the phase 'a' fundamental waveform at site 3 (commercial)

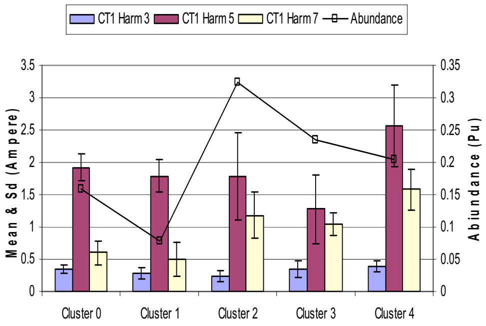

Figure 5(b): Abundance, mean and standard deviation for each cluster of harmonic currents (3rd, 5th and 7th) on phase 'a' at commercial site

Figure 6(a) shows the clusters obtained from substation site (site 1) superimposed on the fundamental current measurement data for two days. Figure 6(b) shows the 7th harmonic current and 7th harmonic voltage at the substation. Figure 6(c) shows the MVAr measurement at the 33kV side of the power system. From Figures 6(b) and (c), it can be observed that the second cluster (s2) is related to the capacitor switching event. Early in the morning, when the system MVAr is high as shown in Figure 6(c), the capacitor is switched on in the 33kV side to reduce bus voltage and late at night when the system MVAr is low, the capacitor is switched off to avoid excessive voltage rise.

The switching-on and switching-off of the capacitor is clearly reflected in the 7th harmonic current as shown in Figure 6(b). The capacitor switching operation in the 33KV side can also be detected at the other sites (residential, commercial and Industrial). Figure 7 shows the case for the commercial site (site 3), where the effect of the capacitor switching operation can be easily observed.

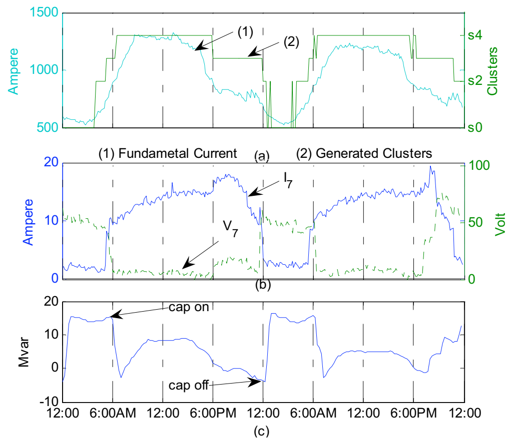

Figure 6: Clusters at substation site in two working days (a) Clusters superimposed on the fundamental current waveform,. (b) 7th harmonic current and voltage data. (c) MVAr load at the 33kv busbar.

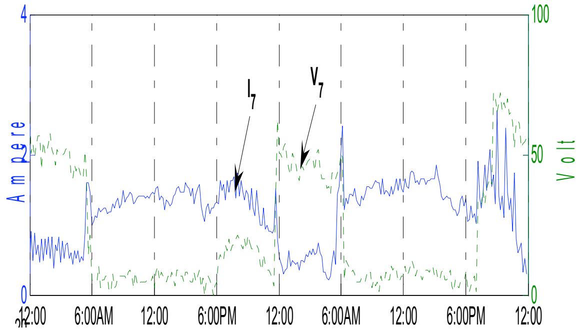

Figure 7: 7th harmonic current and voltage at Commercial site in two working days

The third cluster (s3) appears to occur in the evening period, when Figure 6(b) shows that the 7th harmonic currents rise around 6pm. and drops off around 11 pm. Upon discussion with the Utility Engineer, we were informed that this may be due to the reduction of industrial load and hence the percentage of the harmonics against the fundamental current becomes higher. It was also suggested that this could be due to the switching-on of TVs and appliances when people come home from work.

The same cluster is now applied to a week data from the residential site (site 2). The fundamental current during that week and the temperature data are shown in Figures 8(a) and (b). In this particular week, the Monday temperature (day 3) is relatively high, showing high fundamental current in day 3, resulting in significant use of air conditioners. Also in day 6, the fundamental current can be observed to increase significantly. To analyse these data, initially we concentrate looking at day 3 and 4 and observe how the data mining program cluster the data in these two days. This is then followed by the observation of the day 6 data to see if the data mining program can discriminate the high current data accurately.

Figure 9(a) shows the clusters superimposed on the fundamental current data of days 3 and 4. Cluster s0 is off peak and cluster s3 is on peak. Cluster 1 is representing the switching on and off of the capacitor at the 33KV side. Cluster s2 is representing turning on of TV's and any switch mode power supply (SMPS), in which period the 5th and 7th harmonic currents tend to be high (Fig 9-b). The reason this cluster start early in the morning rather than afternoon is because of the school holiday.

From Figure 9(a) it can be observed that there is a period of peak load (cluster s3) around midnight, and following discussion with the utility engineer, we were told that this is related to the turning-on of off-peak water heaters. Another on peak load (cluster s3) is from 11:00am to 8:00 pm, and from the temperature measurements (Figure 8-b) near the substation area, it can be concluded that air conditioners are the suggested load at that time.

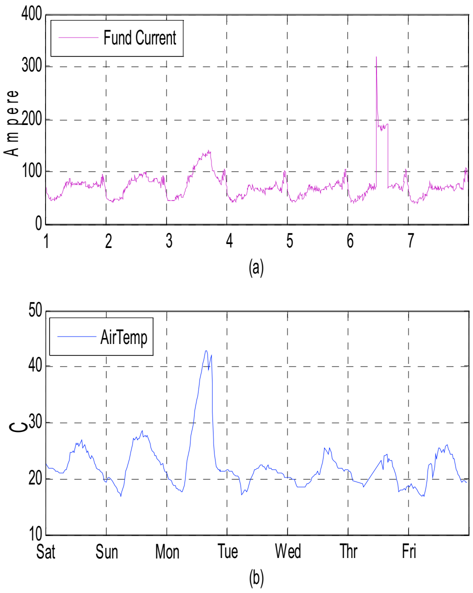

Figure 8: Fundamental current and air temperature in the area of the substation under study for one week from 13/01/2001 to 19/01/2001

Figure 10 shows the clusters obtained for day 6 when the fundamental load current rises in an unusual manner. It can be observed that the data mining program accurately identifies the abnormal event as a separate cluster (cluster 4) between 11:30 am until 4:30 pm in Figure 10(a). Further discussions with the utility engineer, revealed that this is a load transfer event from a faulty feeder to the residential feeder at that time. 5th and 7th harmonic currents are found to be high at that time.

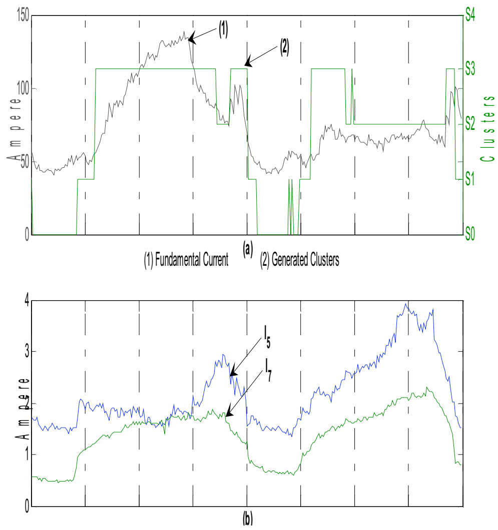

Figure 9: Clusters at residential site for two days (a) Clusters superimposed on the fundamental current waveform, (b) 7th harmonic current and voltage data

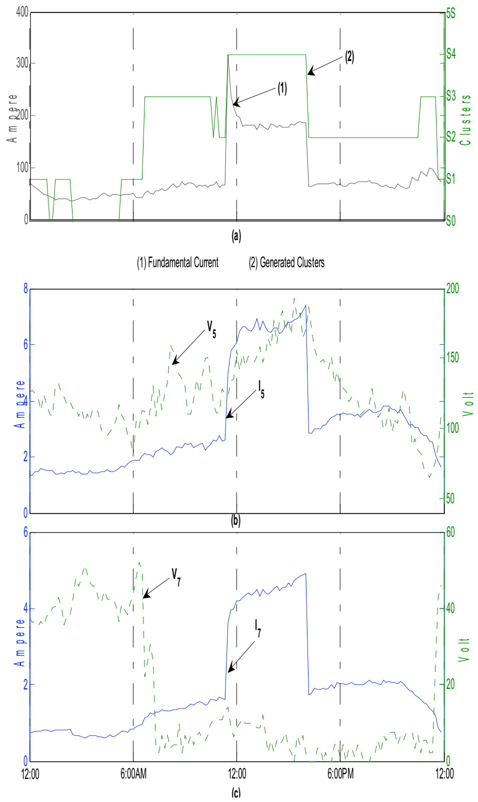

Fig.10: Overload on residential feeder due to transferring load from faulty feeder (a) Clusters superimposed on the fundamental current waveform, (b) 5th harmonic current and voltage data, (c) 7th harmonic current and voltage data

6 Conclusion

Power quality (PQ) data from a harmonic monitoring program in an Australian MV distribution system containing residential, commercial, and industrial customers has been analyzed using data mining techniques. Data mining, and in particular cluster analysis has been shown to be able to identify useful patterns within the data set. The utility engineers as experts in the domain can make ready use of the clustered data to quickly interpret these patterns, and in particular to detect unusual power quality events. The availability of this rapid, visually oriented method of evaluating the underlying information contained in large data sets can be an invaluable tool for power utility engineers.

7 References

Carpinelli, G., P. Caramia, et al. (1996). Probabilistic evaluation of the economical damage due to harmonic losses in industrial energy system. *Power Delivery, IEEE Transactions on* **11**(2): 1021-1031.

Cheeseman, P. and J. Stutz (1995). Bayesian Classification (AUTOCLASS): Theory and Results. In *Advances in Knowledge Discovery and Data Mining.* U. Fayyad, G. Piatetsky-Shapiro, P. Smyth and R. Uthurusanny. Menlo Park, AAAI press**:** 154-180.

Duggan, E. and R. E. Morrison (1993). Prediction of harmonic voltage distortion when a nonlinear load is connected to an already distorted supply. *IEEE* **40**(3): 161-165.

EDMI (2000). Users Manual - EDMI 2000-04XX Energy Meter, Electronic Design and Manufacturing International.

Emanuel, A. E. and et al (1993). A Survey of Harmonic Voltages and Currents in Customer's bus. *IEEE Trans. Power Delivery* **8**(1): 411-421.

Gosbell, V., D. Mannix, et al. (2001). Harmonic Survey of an MV distribution system. *AUPEC, 23-26 September 2001*, Perth.338-342

Heydt, G. T. (1998). Electric Power Quality: A Tutorial Introduction. *IEEE Computer Applications in Power* **11**: 15-19.

Khan, A. K. (2001). Monitoring Power for the future. *IEEE Power Engineering Review Journal* **15**(2): 81-85.

Lachaume, J., T. Deflandre, et al. (1993). Harmonics in MV and LV distribution systems - Present and Future Levels. *IEE Disturbances and protection in Supply Systems*.

M. A.T. Figueiredo and A. K. Jain (2002). Unsupervised Learning of Finite Mixture Models. *IEEE Transactions on Pattern Analysis and Machine Learning* **24**(3): 381-396.

McGranagahn, M. (2001). Trends in Power Quality Monitoring. *IEEE Power Engineering Review Journal* **21**(10): 3-9, 21.

McLachlan, G. (1992). Discriminant Analysis and Statistical Pattern Recognition. New York, Wiley.

Morrison, R. E. and E. Duggan (1993). Long Term Monitoring of power system harmonics. *IEEE Colloquium (Digest)*(120): 2/1-2/3.

Oliver, J., R. Baxter, et al. (1996). Unsupervised Learning using MML, in "Machine Learning. *13th Int. Conf ICML '96*.364-372

Oliver, J., T. Roush, et al. (1998). Analysis Rock Samples for the Mars Lander. *American Association for Artificial Intelligence*: 299-303.

R. Xu and D. Wunsch II (2005). Survey of Clustering Algorithms. *IEEE Transactions on Neural Networks* **16**(3): 645-678.

Robinson, D. (2003). Harmonic Management of MV Distribution Systems. PhD Thesis, School of Elec., Comp. and Telecom. Engineering, University of Wollongong.

Sabin, D. D., D. L. Brooks, et al. (1999). Indices for assessing harmonic Distortion from Power Quality Measurements: Definitions and benchmark Data. *IEEE Trans. Power Delivery* **14**(2): 489-496.

Shuter, T. C., H. T. Vollkommer, et al. (1989). Survey of harmonic levels on the American Electric Power Distribution System. *IEEE Trans. Power Delivery* **4**(4): . 2204-2213.

Shwehdi, M. H., A. H. Mantawy, et al. (2002). Harmonic flow study and evaluation of a petrochemical plant in Saudi Arabia. *Power Engineering 2002 Large Engineering Systems Conference on, LESCOPE 02*.165-172

Stirling, D. and P. Zulli (2004). Ontology Trend Analysis of Dynamic Signals. *International Conference on Intelligent Sensors, Sensor Networks and Information Processing, ISSNIP2004*, Melbourne, Australia.445-449

Wallace, C. (1968). Intrinsic Classification of Spatially Correlated Data. *The Computer Journal* **41**(8).

Wallace, C. and D. Dowe (1994). Intrinsic classification by MML - the snob program. *Proc. 7th Australian Joint conf. on Artificial intellegence*, Armidale, Australia, World Scientific Publishing Co