1-1-2008

Power quality survey factor analysis using multivariable linear regression (MVLR)

Chandana Herath
University of Wollongong

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

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

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

Power quality survey factor analysis using multivariable linear regression (MVLR)

Abstract

During the past two decades, there has been a considerable number of Power Quality (PQ) monitoring programs completed throughout the world. The information collected during these surveys can provide a detailed picture of the expected electrical environment help utilities to plan their future networks in relation to power quality performance. The mass of data gathered for a sample of sites of a large-scale power quality (PQ) survey of this nature has the potential to reveal good and bad influences on power quality if an appropriate procedure for analysis can be determined. If it is known which characteristics are more important in determining the levels of particular PQ disturbances, it could be expected that other sites with similar characteristics would also present similar PQ levels. This paper aims to assess the levels of PQ disturbances and factors that influence good or bad PQ levels in a large scale PQ survey. For this, the Australian National Power Quality Benchmark Survey data has been analysed and compared for the dominant factors in long term PQ data. In this analysis, Multivariable Linear Regression (MVLR) has been identified as a useful tool for factor analysis of complex power quality data.

Keywords

Power, quality, survey, factor, analysis, using, multivariable, linear, regression, MVLR

Disciplines

Physical Sciences and Mathematics

Publication Details

C. Herath, V. J. Gosbell, S. Perera & D. A. Stirling, "Power quality survey factor analysis using multivariable linear regression (MVLR)," in ICHQP 2008: 13th International Conference on Harmonics & Quality of Power, 2008, p. [5].

Power Quality Survey Factor Analysis using Multivariable Linear Regression (MVLR)

H.M.S.Chandana Herath, *Student Member, IEEE,* V.J. Gosbell, *Member, IEEE,* S. Perera, *Member, IEEE,* and D. Stirling, *Member, IEEE*

***Abstract*- During the past two decades, there has been a considerable number of Power Quality (PQ) monitoring programs completed throughout the world. The information collected during these surveys can provide a detailed picture of the expected electrical environment help utilities to plan their future networks in relation to power quality performance. The mass of data gathered for a sample of sites of a large-scale power quality (PQ) survey of this nature has the potential to reveal good and bad influences on power quality if an appropriate procedure for analysis can be determined. If it is known which characteristics are more important in determining the levels of particular PQ disturbances, it could be expected that other sites with similar characteristics would also present similar PQ levels. This paper aims to assess the levels of PQ disturbances and factors that influence good or bad PQ levels in a large scale PQ survey. For this, the Australian National Power Quality Benchmark Survey data has been analysed and compared for the dominant factors in long term PQ data. In this analysis, Multivariable Linear Regression (MVLR) has been identified as a useful tool for factor analysis of complex power quality data.**

***Index Terms*- Factor analysis, Power Quality (PQ) data analysis, PQ monitoring.**

I. Introduction

The mass of data gathered for a sample of sites of a large-scale power quality (PQ) survey has the potential to reveal good and bad influences on power quality if an appropriate diagnostic procedure can be determined. Survey results from a particular set of monitored sites can be used to infer the PQ behaviour of other unmonitored sites with similar characteristics in terms of physical behaviour and similar mix of customers. This would be useful for network planners for network planning and other related power quality studies. This is only possible if the characteristics that are the most influential in determining PQ levels at a particular site are known.

Factor analysis can be used to enable the factors which contribute to poor or better PQ levels to be identified. Results of such an analysis are used in PQ management, planning and reporting practices in utility environments.

This paper aims to assess the levels of PQ disturbances and factors that influence good or bad PQ levels in those 70 sites of 9 utilities in Australia [1]. The characteristics of the sites that were monitored, the methodology adopted on the survey measurements in connection with the site selection, location of PQ monitors, site types and load types that have been used to identify different mix of customers within variety of load types will be detailed. The survey data has been analysed using Multivariable Linear Regression (MVLR).

Based on the factor analytic models developed in this paper, it is determined whether any relationship exists between those categorical variables (i.e. site types and load types) and PQ disturbance levels measured in the survey.

II. Discussion on the Available Survey Data

After an extensive search for PQ data analysis techniques for factor analysis, it was shown that Multivariable Linear Regression (MVLR) is a reliable and useful analysis tool for PQ survey data analysis. It is shown that MVLR allows insights to be developed without having the types of models required for factor analysis

A. Site Selection

The measurement locations in the 240/415V networks were selected within each participating utility's network in consultation with the utility. The sites were spread across geographic areas and bulk supply points, and across load categories, i.e. sites that were predominantly residential, commercial, industrial, rural, or remote. One set of site locations was selected in places likely to show power quality problems. The remaining set of site locations were selected from more normal parts of the network. Common site selection criteria were applied in all cases. During the selection process, the following were considered:

(i) Fault history of the network
(ii) Operating diagrams/maps showing feeder lengths and circuit arrangements
(iii) Information on loads which have a major impact on network
(iv) Relevant previous survey data
(v) Major planned system alterations over the period of the survey.

This work was supported by the Integral Energy Australia under an Australian Research Council (ARC) Grant.

H.M.S.C. Herath is with Energy Australia and a Post graduate Student at University of Wollongong. NSW 2522, Australia.

V.J. Gosbell, S. Perera and D. Stirling are with the School of Electrical Computer & Telecommunications Engineering, University of Wollongong, NSW2522, Australia (Corresponding author's E-mail: CHerath@energy.com.au).

978-1-4244-1770-4/08/$25.00 ©2008 IEEE

B. Defining Categorical Variables

Any variable that is not quantitative is categorical. Categorical variables take a value that is one of several possible categories [2]. Defining categorical variables in this study, (i.e. Site and Load categories) were based on the information obtained from each utility. When analysing categorical data, we will typically work with *counts* or *percentages* of objects (i.e. site or load type) which fall within certain categories. Once the load category is defined as Residential, Commercial, Industrial, Rural and/or Remote, site categories were identified as (a) Poor sites, i.e. Extreme sites were usually selected at the end of the 240/415V distributor (feeder) (b) and the Average sites were usually located at the beginning of the feeder or a distributor close to the 11kV/ 415V Substation. Following guidelines were followed in the selection of sites with particular power quality problems:

(i) Steady state voltage and unbalance – residential areas on outskirts of town
(ii) Sags – industrial areas that lightning prone or an areas which may see the impact of motor starting
(iii) Harmonics – commercial areas (eg. shopping complexes) outside of city centre
(iv) Transients – areas affected by capacitor switching and lightning

The sites chosen were preferably on mixed distributors (residential, commercial, industrial) even though one load type may have dominated. Another preference was to nominate sites with the possibility of being dominated by each load category.

III. MULTIVARIABLE LINEAR REGRESSION (MVLR)

A. General Introduction

Multivariable Linear Regression (MVLR) [3] enables fitting of a single model for categorical variables such as site type and load type, taking into account the relationships among those power quality data categories of a single disturbance type of many sites. The factor analysis of PQ data can allow uneven representation of different factors, which can be used to establish useful insights on monitored PQ of all sites. In this paper, the MVLR analysis is carried out using the MSExcel, Solver add in with least squares fit.

The site categories used in MVLR are:

- Site type (Average, Extreme)
- Load type (Residential, Commercial, Industrial, Rural, Remote)

It is important to determine whether site type or load type is the most important factor determining a PQ problem of interest.

For example, assume that the Voltage Unbalance Factor (VUF) is related to the categorical variables as given by (1)

$$VUF = K_0 + K_{Site}(Site\,type) + K_{Load}(Load\,type) \quad (1)$$

where, the constants K_{Site} and K_{Load} are to be determined by means of Multivariable Linear Regression. To reduce the number of variables, it is assumed that $K_{Site}(Average)$ and $K_{Load}(Commercial)$ are equal zero, i.e. they are incorporated into K_0 . Equation (1) is solved for all constants using a least squares fit in MS Excel solver. Description of the methods are described in the section below.

B. Example demonstrating the application of MVLR

Suppose, each site is categorised by,

- Site type (Average, Extreme)
- Load type (Commercial, Industrial)

Also, suppose there is a need to determine whether site type or load type is the most important factor determining voltage unbalance.

Assume, Voltage unbalance factor (VUF),

$$VUF = K_0 + K_{Site}\,(Com, Ind) + K_{Load}\,(Av, Ext)$$

where, each constant to is be determined by means of Multivariable Linear Regression.

To reduce the number of variables, it is assumed that K_{Site} (Average) and K_{Load} (Commercial) are zero, i.e. they are incorporated in K_0. It is solved for all constants using a least squares fit in MS Excel solver. The two examples given below give a better representation of the above methodology.

Example 1: This example demostrates as to how the constants are incorporated into K_0 when they are forced to zero.

(i) VUFs with no constant (K_0) introduced:

	Site type		Load type	
	Suburban	CBD	Commercial	Industrial
VUF	0.2	0.3	0.2	0.1

Based n the above,

(a) Suburban commercial sites will have a VUF of 0.4 (i.e. 0.2 + 0.2).
(b) CBD commercial will have a VUF, 0.5 (i.e. 0.3+0.2).
(c) Suburban industrial will have a VUF, 0.3 (i.e.0.2 + 0.1).
(d) CBD industrial will have a VUF, 0.4 (i.e. 0.3+0.1)

(ii) VUFs with constant (K_0):

In this section the constant (K_0) will introduce by forcing some indices to zero such that the configuration of above data will have no effect as below:

		Site type		Load type	
	K_0	Suburban	CBD	Commercial	Industrial
VUF	0.4	0	0.1	0	-0.1

After introducing the constant (K_0)above indices will calculate as below,

(a) Suburban commercial sites will have a VUF of 0.4 (i.e. 0.4 + 0 +0).
(b) CBD commercial will have a VUF, 0.5 (i.e. 0.4 + 0.1 + 0).
(c) Suburban industrial will have a VUF, 0.3 (i.e.0.4 + 0 - 0.1).
(d) CBD industrial will have a VUF, 0.4 (i.e. 0.4 + 0.1 –0.1).

Example 2: This example gives an analysis of voltage unbalance based on the regression equation that has been constructed using set of voltage unbalance data of a number of sites for a situation similar to Example 1.

The analysis of 100 sites shows that the unbalance can be approximated reasonably well by the (2).

$$\text{VUF} = 0.5 + [0.3\ (\text{Suburban}) + 0.1\ (\text{CBD})] + 0.1(\text{Commercial}) + 0.2\ (\text{Industrial})] \quad (2)$$

Based on (2) many predictions can be made in relation to the voltage unbalance, some of which are given below:

(a) Predicted VUF in relation to commercial loads of suburban sites is,
0.5+0.3+0.1 = 0.9
(b) Most important factor determining unbalance is suburban (contributes to range of 0.3)
(c) Lowest levels of unbalance are found at commercial locations of CBD sites.

IV. Factor Analysis of Australian PQ Data

A. Overview

Based on the methodology described above, survey data [1] has been analysed to describe a factor analysis model representing the Australian utilits. The details of the survey including the types of disturbances monitored based on the site selection described above, are discussed further.

The measured quantities have been divided into two groups, namely, those that were measured by logging and those that were captured as distinct events. The power quality survey was carried out using the following voltage quantities at each site for approximately one week period.

Logged quantities (Continuous disturbances)

1. Voltage Unbalance
2. Voltage harmonics
3. Steady State Voltage

Normalised indices [4] have been calculated for each site by analysing the survey data for continuous disturbances. Indices for captured quantities (discrete disturbances) have been left out as they were not measured in all sites.

B. Relationships of factors on Individual Disturbance Indices

The analysis given below is established separately for each individual disturbance by means of the factors relating to Site and Load types. Site type mainly describes the impedance characteristic of each site whereas the Load type is mainly based on how much current is flowing to the load depending on each load category.

The regression equation by means of MVLR analysis for all the types of disturbance indices is given by,

Disturbance Index

$$= K_0 + K_{Sitype}(Av, Ex) + K_{Ltype}(Rs, ..) \quad (3)$$

In this analysis, Commercial and Average site indices were forced to zero, i.e. they are incorporated into K_0 in MVLR analysis for all disturbance types described below.

C. Voltage Unbalance

In Table I, it was noted that there were missing data for unbalance measurements in remote sites. This is due to a fault in the PQ monitors that have been downloading data for those sites. Therefore, the Remote sites were excluded from the unbalance analysis. In general Residential, Rural and Remote sites would show a significant Voltage Unbalance Factors (VUFs) compared to Commercial and Industrial sites. This is mainly because - there are more single phase unbalanced loads present in residential areas and more scattered and isolated unbalanced loads are present in rural and remote areas.

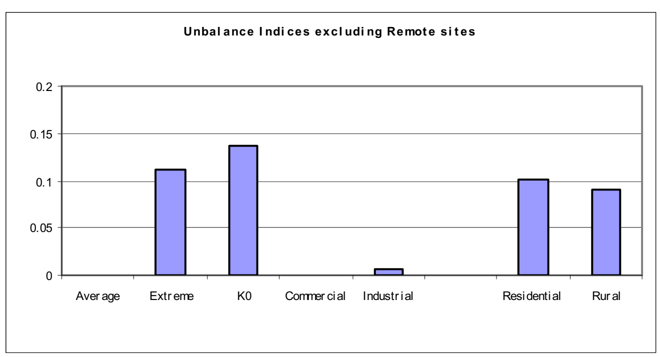

Fig.1 MVLR Analysis on Voltage Unbalance

TABLE I
MVLR ANALYSIS ON VOLTAGE UNBALANCE

Constant	Site Type		Load Type			
K_0	Average	Extreme	Commercial	Industrial	Residential	Rural
0.136	0	0.111	0	0.005	0.101	0.091

MVLR analysis shows that the load category contributes to significant unbalance levels compared to the site category. Fig 1 also shows that there is little difference at average or extreme sites in terms of unbalance levels. There is no significant contribution from site category for determining the voltage unbalance. K_0 (constant) and Extreme values do not have a much difference between them (K_0 and Extreme 0.136 and 0.111 respectively). When compared with the variation of indices, the load type has a significant input and large variation in its components for determining the poor or better performance in voltage unbalance levels. Therefore, in conclusion the Site type is not a significant factor for deciding good or bad voltage unbalance levels.

Calculation of the contribution of Voltage Unbalance Factor (VUF) for site types in relation to Commercial & Industrial load types,

$$\begin{aligned} \text{VUF (Ave, Com)} &= \text{K0} + \text{KS (Average)} + \text{KL (Commercial)} \\ &= 0.136 + 0 + 0 \\ &= 0.136 \end{aligned}$$

Calculation in relation to VUF shows that Unbalance contributions of Industrial & Commercial sites are slightly similar. Values for Average & Extreme sites for Commercial & Industrial are 0.136, 0.141 and 0.247, 0.252 respectively. The Extreme sites i.e. the sites further away from the transformer have higher contribution to the Voltage unbalance as expected in general scenarios. Also, Figure 1 shows that Unbalance contribution of Residential and Rural load categories is similar. However, the worst unbalance levels are reported Extreme/Residential sites. The most significant factor for Voltage Unbalance is Extreme/ Residential, followed by Extreme/Rural. The Extreme sites show a significant difference with Average sites, i.e. a factor of 0.111 difference between the site categories.

TABLE II
FACTORS CONTRIBUTING TO VOLTAGE UNBALANCE

Description	VUF Contribution
VUF (Average, Commercial)	0.136
VUF (Average, Industrial)	0.141
VUF(Extreme, Commercial)	0.247
VUF(Extreme, Industrial)	0.252

It is to be noted that the exclusion of remote sites in MVLR analysis described above have only been considered for Voltage Unbalance. As discussed above, it is because that the Unbalance data for Remote sites were not recorded by PQ monitors of those sites. However this is not the case for Voltage harmonics and Steady state voltage.

D. Voltage Harmonics

Harmonic data of the survey were logged for all 70 sites and no data loss has been reported. Therefore the MVLR analysis on Harmonic has included in all the sites measured in the survey. Also in this analysis, the Commercial and Average site indices were forced to zero for consistency where those values are incorporated in K_0.

Results of MVLR analysis illustrated in Fig 2 show that Extreme sites have worst Harmonic levels in the system; in particular Extreme/Commercial are the worst performing sites for Harmonics, followed by Extreme/Residential, Rural, Industrial and Remote. Similar to the analysis for Unbalance and Voltage, the Commercial and Industrial sites were compared for Voltage Harmonics as shown below.

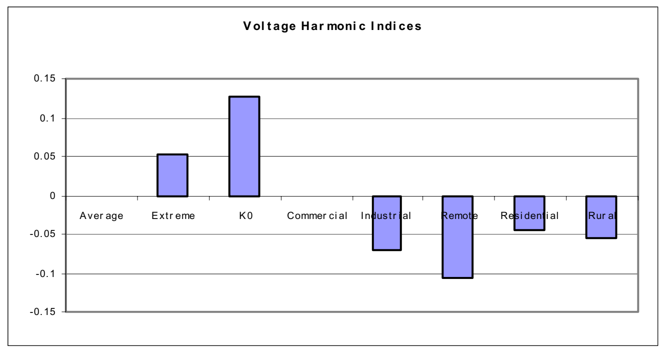

Fig.2 MVLR Analysis on Voltage Harmonics

TABLE III
FACTORS CONTRIBUTING TO VOLTAGE HARMONICS

Constant	Site Type		Load Type				
K_0	Average	Extreme	Commercial	Industrial	Remote	Residential	Rural
0.128	0	0.054	0	-0.072	-0.107	-0.046	-0.055

Analysis shows that the Commercial sites are the worst performing for harmonics in relation to load category. This is followed by Residential, Rural, Industrial and Remote. Also, the Extreme sites i.e. sites further away from the transformer contribute more Harmonics into the system than Average sites i.e. those sites closer to the transformer. Therefore, it is noted that both site type and load type are contributing factors for harmonic performance of a utility network where Extreme/ Commercial sites are the worst performing ones for harmonics.

TABLE IV
FACTORS CONTRIBUTING TO VOLTAGE HARMONICS (VHARM)

Description	VHarm Contribution
VHarm (Average, Commercial)	0.128
VHarm (Average, Industrial)	0.056
VHarm (Extreme, Commercial)	0.182
VHarm (Extreme, Industrial)	0.110

E. Steady State Voltage

Similar to the analysis of voltage unbalance and harmonics in MVLR, the indices for Average and Commercial site were forced to zero for consistency where those indices were incorporated in K_0(constant). The difference in the voltage analysis compared to unbalance has shown a significant variation between K_0 and site and load categories. Both Site type and Load type are significant factors for deciding good or bad steady state voltage levels. Therefore, the factor of being close to the transformer or further away from the transformer is one of the deciding factors for steady state voltage.

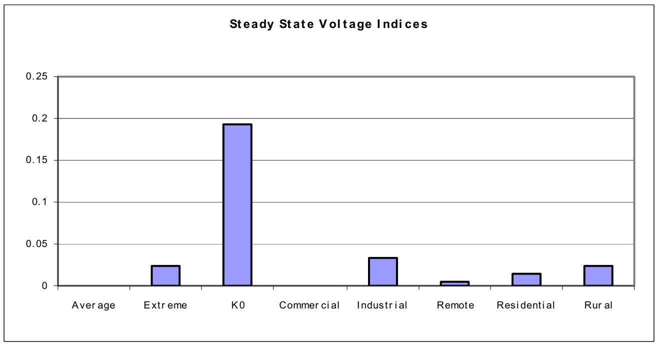

Fig.3 MVLR Analysis on Steady State Voltage

TABLE V
MVLR ANALYSIS ON STEADY STATE VOLTAGE

Constant	Site Type		Load Type				
K_0	Average	Extreme	Commercial	Industrial	Remote	Residential	Rural
0.192	0	0.025	0	0.033	0.005	0.013	0.024

As for voltage unbalance and harmonics, the Commercial and Industrial sites data were compared in relation to Steady State Voltage as given in Table VI.

TABLE VI
FACTORS CONTRIBUTING TO STEADY STATE VOLTAGE (SSV)

Description	SSV Contribution
SSV (Average, Commercial)	0.192
SSV (Average, Industrial)	0.225
SSV (Extreme, Commercial)	0.217
SSV (Extreme, Industrial)	0.250

As expected, the Extreme sites are generally the worst performing sites for steady state voltage. The MVLR analysis shows that the most contributing factor for voltage is Extreme/Industrial sites, followed by Extreme/Rural, Extreme/Residential, Extreme/Remote and Extreme/Commercial. However, it is also to be noted that there is no significant difference between Extreme and Average sites, i.e. 0.025. Industrial sites are the most contributing factor for Steady State Voltage followed by Rural, Residential, Remote and Commercial sites.

V. CONCLUSIONS

This paper discussed the application of factor analysis in relation to a PQ Survey data. Factor analysis has been used to identify hidden patterns and relationships that reveal the factors contributing to poor or good power quality performance in a utility network. For this analysis, a sample of Australian PQ survey data has been analysed and compared for the dominant factors. Multivariable Linear Regression (MVLR) has been identified as a useful tool for factor analysis of complex power quality data.

In the analysis considered there are two major factors considered based on the survey data available, i.e. Site type and Load type. Site type mainly describes the impedance characteristics whereas Average sites are those sites close to the Transformer and Extreme sites are the ones further away from the transformer. Load type is related to the type of current disturbances which are present in the load and are identified as Commercial, Industrial, Residential, Rural and Remote load categories.

The study found that the most significant factor contributing to PQ levels is the load type, i.e. for unbalance, harmonics and steady state voltage. However, both load type and site type contribute to the steady state voltage.

It is shown that the factor analysis is a specific analysis technique that can be used to identify hidden patterns and relationships which can be accomplished using MVLR.

Further research should be aimed at using Data Mining Techniques for factor Analysis and compare both MVLR and Data mining techniques to come up with a solution to determine which method is more suitable for such analysis.

REFERENCES

[1] Gosbell, V.J., Perera, S., Smith, V.W., *National Australian Benchmark Power Quality survey Report.* Integral Energy Power Quality Centre, University of Wollongong, NSW, Australia, 2001.

[2] A. Agresti, "Categorical data analysis", Wiley-Interscience, New York, 2002.

[3] H.M.S. Chandana Herath, PhD Thesis "Power Quality Data Management and Reporting Methodologies for Australian Electricity Distribution Industry", University of Wollongong, NSW, Australia, (Pending Submission) 2008.

[4] Gosbell, V.J., Perera, S., Herath, H.M.S.C., Unified Power Quality Index (UPQI) for continuous disturbances. 10th IEEE International Conference on Harmonics and Quality of Power (ICHQP'02), Rio de Janeiro, Brazil, October 2002, Vol. 1, pp. 316 – 321.

BIOGRAPHIES

Chandana Herath (S'2000): Mr. Herath is currently working as an Engineer at Energy Australia (EA) and also a post graduate student studying towards his PhD in Power Quality, at the Integral Energy Power Quality Centre (IEPQRC) at the University of Wollongong, Australia. He is representing EA for ENA Reliability & Power Quality Working Group and EL-034 Standards Australia PQ Committee.

Vic Gosbell (M'1972): Dr. Gosbell is the former Technical Director and current Technical Advisor to the IEPQRC and former Professor of Power Engineering, University of Wollongong, Australia. He is a fellow of the Institution of Engineers, Australia and currently a Professorial Fellow in IEPQRC at the University of Wollongong. His research interests are in Power Electronic Simulation, Harmonics and Power Quality.

Sarath Perera (M'1988): Dr. Perera an Associate Professor at the University of Wollongong, Australia and is the Technical Director of IEPQRC. His research interests are in Power Quality.

David Stirling (M'2001): Dr. Stirling is a Senior Lecturer at the University of Wollongong, Australia. His research interests are Machine Learning and Data Mining.