Voltage Total Harmonic Distortion Analysis through Transformer's Characteristics and Energy Use Data

W. E. Souza, A. R. Aoki, A. Chaves Neto, R. A. Peniche and M. R. Ortega

***Abstract*—This paper describes the statistical analysis of a harmonic monitoring campaign on distribution transformers as well as the development of a computational statistical model based on multivariate analysis techniques to predict voltage total harmonic distortion (THD) through transformer's characteristics and energy use data.**

***Keywords:* Voltage Total Harmonic Distortion, Logistic Regression, Quadratic Score.**

I. Introduction

IN the last years, Power Quality (PQ) had increasing interest to Power System's Agencies due to decreasing reliability and service indices. In other side, equipment evolution more sensitive to power disturbances, consumer requirements, deficit of standards specifying PQ limits, has contributing with this problem. Moreover, the grow cost of equipment repair because of damage caused by disturbances could be reverted in power systems investments.

The consumer loads in distribution systems have even more non linear characteristics inducing voltage harmonic distortions. This kind of pollutant loads, non linear loads, expose consumers and utilities to raised harmonic distortions producing unwanted effects like losses increasing, equipment malfunction, and harmonic over voltages. The phenomenon has higher relevance if it's considered the problem of assigning responsibilities between consumers and supplier about harmonic injection on distribution networks.

This work shows a statistical analysis of a harmonic monitoring campaign on distribution transformers and establishes correlation between transformer's technical characteristics and consumer's energy use data to voltage THD [1] in low voltage distribution network. To achieve this purpose we used statistical models, based on multivariate analysis [2], and COPEL (Paraná Energy Company – Brazil) transformers data with its harmonic measurements.

This work was supported in part by the Paraná Energy Company (COPEL).
W. E. Souza and A. R. Aoki are with Institute of Technology for Development - LACTEC, Centro Politécnico da UFPR, PO Box 19067, Zip Code 81590-980, Curitiba/PR – Brazil, Phone +55 41 3361-6012 or Fax +55 41 3361-6007 (e-mail of corresponding author: aoki@lactec.org.br).
A. Chaves Neto is with Paraná Federal University, Centro Politécnico da UFPR, PO Box 19081, Zip Code 81531-990, Curitiba/PR - Brazil (e-mail: anselmo@ufpr.br).
R. A. Peniche and M. R. Ortega are with Paraná Energy Company - COPEL, Rua Padre Agostinho, 2600 Zip Code 80710-000, Curitiba/PR - Brazil (e-mail: rodrigo.peniche@copel.com).

Paper submitted to the International Conference on Power Systems Transients (IPST2009) in Kyoto, Japan June 3-6, 2009

II. Power System Voltage Harmonic Monitoring

The database, used on classification models, was obtained through a sampling monitoring campaign covering all distribution transformer types operating in COPEL power system. The monitoring campaign has the intention of mapping voltage harmonic distortion levels in COPEL distribution system.

It was monitored 399 transformers in 5 regions of COPEL permission area. These equipments were selected in an universe of about 320,000 distribution transformers considering regional installed transformer capacity and the number of consumers by transformer in urban and rural areas.

Table I shows the summary of only urban transformers measured.

TABLE I
MONITORING DISTRIBUTION ACCORDING TO REGIONAL INSTALLED POWER

Region	Location	Samples
SDC	Center	32
SDL	East	128
SDN	Northwest	63
SDO	West	62
SDT	North	74
TOTAL		361

The monitoring equipment parameterization and monitoring time for each distribution transformer were established by Brazilian Electricity Regulatory Agency (ANEEL) resolution 505 considering 10 minutes for integration time during 7 consecutive days [4].

Voltage THD maximum, minimum and mean levels, as well as, standard deviation and $P_{95\%}$ was analyzed for each monitored transformer. Although for statistical model we used only $P_{95\%}$ of voltage THD with distribution transformers technical characteristics and energy use data. It was used a Brazilian regulation to evaluate the maximum value of voltage THD and voltage levels allowed in low voltage distribution networks [3]-[4].

A. *Transformers Technical Characteristics and Energy Use Data*

Based on monitoring results it was looked for power network parameters that can affect voltage harmonic distortion behavior in low voltage distribution networks to provide statistical models development. So, it was considered as determining factor to voltage waveform degradation in low voltage distribution networks the load characteristics and short-circuit power in the connection point.

The load characteristics were divided into monthly mean energy use by each consumer class connected in the distribution transformer. Moreover, characteristics of the equipment were used to enhance the models. Table II shows the characteristics used.

TABLE II
TECHNICAL CHARACTERISTICS AND ENERGY USE DATA

Characteristics	Description
Operation	Urban or rural
Voltage [V]	34,500 or 13,800
Power [kVA]	112.5, 75, 45, 30, 15, 10 or 5
Number of Phases	1 or 3
Number of Consumers	Number of Consumers connected to distribution transformer
Energy [kWh]	Monthly mean energy consumption in each consumer class: • Residential 0 to 50 • Residential 50 to 100 • Residential 100 to 150 • Residential 150 to 200 • Residential 200 to 300 • Residential 200 to 300 • Residential 300 to 500 • Residential 500 to 1,000 • Residential more than 1,000 • Commerce and public power • Industries • Rural • Public illumination

III. STATISTICAL PATTERN RECOGNITION

According to [5], from the statistics point of view, rules development to pattern recognition and classification for two populations is based on four main methods: Fisher linear discriminating function, quadratic discriminating score, logistic regression and k-means method. Moreover, the discriminating analysis is a multivariate technique with the purpose to separate two distinct objects' set and allocates new objects in previously defined sets. When it's used as classification procedure it's not an exploratory technique, since it leads to well defined rules, which can be used for another objects classification.

The use of multivariate techniques to discriminating and classification has the follow intentions.

1. Algebraic or graphic description of object's distinct characteristics from several known populations in order to find discriminating value which leads to the maximum separation of populations.
2. Cluster objects in two or more determined classes trying to find a rule to be used for optimal allocation of a new object. A separation function can be used for object allocation, as well as, an allocating rule can suggest a discriminating procedure.

A. Logistic regression

The technique consists in connect, through a model, the output signal to influence factors of an event. To attend this criteria the multiple linear logistic model must have the following aspect.

Consider a random variable Y (dichotomic) and $\underline{X}'=[X_1, X_2,..., X_p]$ a p dimension vector composed of independents random variables. Taking n independent observations of Y and X_i, with $i=1, 2, ..., p$ we can write logistic regression model like:

$$P(y)=\frac{1}{1+e^{-\mu}} \quad (1)$$

with $\mu=\beta_0+\beta_1 x_1+\ldots+\beta_{p-1}x_{p-1}=\underline{x}\,\underline{\beta}$.

Where y is the dichotomic output signal and it is function of $x_1, x_2, \ldots, x_{p-1}$, $P(y)$ is the probability of occurring y, $x_1, x_2, \ldots, x_{p-1}$ are influence factors, and $\beta_0, \beta_1, \ldots, \beta_{p-1}$ are models parameters.

The function (1), previously defined, has values between 0 to 1 to $\mu \in (-\infty, \infty)$. The figure 1 shows the graphic result of $P(y)$ function.

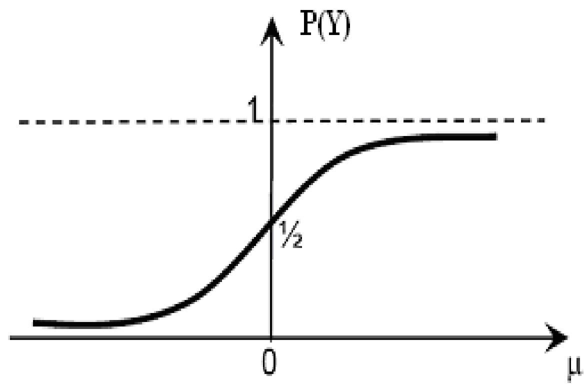

Fig. 1. Graphic result of P(y).

According to [5] this model is the most suitable for dichotomic dependent variable, because logistic regression estimates directly the probability of an event occurrence. Those authors points that parameters estimation must be done by maximum verisimilitude method, this method is the most recommended when the database have individual observations of event occurrence or not.

B. Quadratic discriminating score

Consider the random variable's vector $\underset{\sim}{X}$ from populations with multivariate normal distribution with mean $\underset{\sim}{\mu_i}$ and covariance matrix Σ_i, like,

$$f_i(\underline{X})=\frac{1}{(2\pi)^{p/2}\left|\Sigma_i\right|}\exp\left[-\frac{1}{2}(\underline{X}-\underline{\mu}_i)'\Sigma_i(\underline{X}-\underline{\mu}_i)\right] \quad (2)$$

where $i=1, 2, ..., g$; $\underline{\mu}_i=E(\underline{x}\,|\,\Pi_i)$, and $V(\underline{x}\,|\,\Pi_i)=\Sigma_i$.

Allocate $\underline{X}$ in Π_k if

$$\ln p_i f_i(\underline{X}) = \ln p_i - \left(\frac{p}{2}\right)\ln(2\pi) - \frac{1}{2}\ln|\Sigma_i| - \frac{1}{2}(\underline{x} - \underline{\mu}_i)'\Sigma_i^{-1}(\underline{x} - \underline{\mu}_i) \quad (3)$$

where $i=1, 2, \ldots, g$.

$$ECM = p_1\left(\sum_{k=2}^{g} P(k|1)c(k|1)\right) + \sum_{\substack{k=1\\k\neq 2}}^{g} P(k|2)c(k|2)) + \ldots + \sum_{k=1}^{g-1} P(k|g)c(k|g)) \quad (4)$$

where ECM is the Expected Cost of Misclassification.

In equation (4) the constant $\left(\frac{p}{2}\right)\ln(2\pi)$ can be ignored since is the same for all populations. The quadratic discriminating score is defined as d_i^Q to the population Π_i as:

$$d_i^Q(\underline{X}) = \frac{1}{2}\ln|\Sigma_i| - \frac{1}{2}(\underline{x} - \underline{\mu}_i)'\Sigma_i^{-1}(\underline{x} - \underline{\mu}_i) + \ln p_i,\ i = 1,2,\ldots,g \quad (5)$$

where $i=1, 2, \ldots, g$.

With several normal populations, the classification rule consists in classify $\underset{\sim}{X}$ in Π_k if:

$$d_i^Q(\underline{X}) = \max_i d_i^Q(\underline{X}),\ i = 1,2,\ldots,g \quad (6)$$

where $i=1, 2, \ldots, g$.

IV. Statistical Analysis

The first step was a statistical analysis applied to monitoring campaign database. This was helpful to statistical pattern recognition methods parameterization and transformers clustering.

The database was divided into 5 regions from COPEL distribution systems and Table III shows some statistical results for urban transformers using $P_{95\%}$ of voltage THD.

TABLE III
$P_{95\%}$ OF VOLTAGE THD STATISTICAL ANALYSIS.

Statistics	SDL	SDT	SDC	SDN	SDO
Sample	128	74	32	63	62
Mean	4.8876	4.9006	5.47083	5.00952	4.8973
Variance	4.7103	3.7079	11.8868	10.4524	8.2752
Standard Deviation	2.1703	1.9256	3.44772	3.23301	2.8766
Minimum	2.10	1.80	2.60	1.96	1.99
Maximum	15.89	10.19	16.78	25.18	22.95

Analyzing the mean values presented in Table III of voltage THD obtained in the 5 regions of COPEL power system we can observe that's no apparent difference. Although SDL and SDT have similar mean and variance values, so we can conclude that those regions have analogous distribution systems and load configuration. Another important observation is related to variance values that indicate homogeneity in monitoring results, i.e. transformers outside great urban centers have similar performance to great urban centers' transformers.

The other regions (SDC, SDN and SDO) have similar mean and variance values, therefore the variance values are greater than the values for SDL and SDT regions. The higher variance value indicate a non-homogeneity performance of transformers outside great urban centers when compared to great urban centers' transformers. So these regions determine another electrical characteristics' group.

The worst performance was SDC region showing its power system's fragility to voltage harmonics.

Those conclusions was obtained through a non-parametric test and for variance it was applied the F test.

Considering that most of transformers analyzed by statistical methods have 3 phases, we did a correlation analysis between phases to certify that the maximum $P_{95\%}$ values used was representative for other phases too. Table IV shows phase's correlation values.

TABLE IV
PHASE'S CORRELATION VALUES.

	THDva [%]	THDvb [%]	THDvc [%]
THDva [%]	1	0,817139	0,849125
THDvb [%]	0,817139	1	0,929703
THDvc [%]	0,849125	0,929703	1

We can observe a great correlation between different phases. Figure 2 shows correlation for all considered transformers.

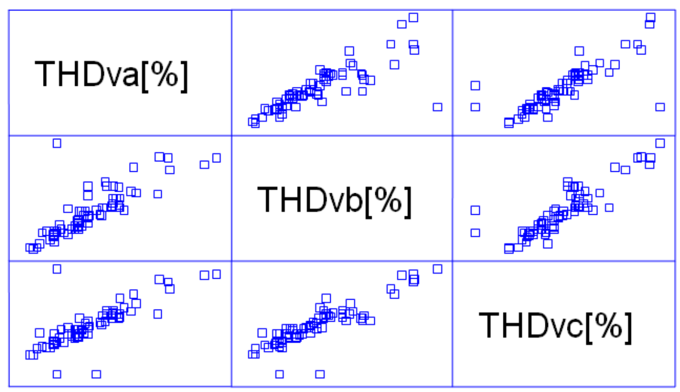

Fig. 2. Phase's correlation for all transformers.

V. Pattern Recognition Methodology

The logistic regression and quadratic discriminating score was used to separate distribution transformers in two groups, one with voltage THD less or equal than 6% and another with voltage THD over than 6% [3].

The limit for voltage THD considered to estimation of statistical models (6%) was obtained from ONS (Power System National Operator) Network Procedures from Brazil. Table V shows limits established by this Institution.

The groups were separated by their maximum value of voltage THD, like:

- Group 1: dependent variable with $P_{95\%}$ of voltage THD over than 6% in any transformer phase;
- Group 2: dependent variable with $P_{95\%}$ of voltage THD less or equal than 6% in all transformer phases.

TABELA V

GLOBAL AND INDIVIDUAL LIMITS FOR VOLTAGE HARMONIC DISTORTIONS (IN PERCENTAGE OF FUNDAMENTAL VOLTAGE)

Voltage < 69 kV			
Odd		Even	
Order	Value (%)	Order	Value (%)
3, 5 and 7	5	2, 4 and 6	2
9, 11 e 13	3	≥ 8	1
15 a 25	2	-	-
≥ 27	1	-	-
Voltage THD = 6%			

The influence factors $x_1, x_2, \ldots, x_{p-1}$ considered to estimation of parameters models were technical data and operational information about distribution transformers monitored, as well as, information about energy use for each consumer class, summing 16 variables.

VI. RESULTS AND DISCUSSIONS

The results from method's application to classification and recognition of distribution transformers voltage THD are shown below.

A. Logistic regression results

The application of Logistic Regression to classification and recognition of distribution transformers voltage THD reached satisfactory results for all regions analyzed. The best result was obtained for SDC (Center-South Distribution Region) with an apparent error rate of 5.88% and the worst result was SDN (Northwest Distribution Region) with an apparent error rate of 25.39% as shown in table VI.

The mean apparent error rate for all regions using Logistic Regression was 14.64%. It was used 0.35 as limit for logistic function, because it's almost the transformer group's proportion and the value with better classification results.

TABLE VI

LOGISTIC REGRESSION CLASSIFICATION RESULTS

Region	Error I	Error II	APER
SDC	8.70%	0.00%	5.88%
SDL	6.80%	16.00%	8.60 %
SDN	23.91%	29.39%	25.39%
SDO	13.73%	27.27%	19.23%
SDT	13.79%	37.50%	18.92%

where:

- Error I: transformers from group 2 classified as group 1;
- Error II: transformers from group 1 classified as group 2;
- APER: apparent error rate.

B. Quadratic discriminating score with equal probabilities results

The application of Quadratic discriminating score with equal probabilities for recognition and classification of distribution transformers voltage THD reached satisfactory results for four regions. The West Distribution Region (SDO) hasn't a satisfactory result, and it was because of matrix bad conditioning of group's database for this region. The database can be considered a sparse matrix, with several null values, causing a non satisfactory method's performance.

The best classification result was obtained for North Distribution Region (SDT) with an apparent error rate of 6.75% and the worst result was West Distribution Region (SDO) with an apparent error rate of 48.39% as shown in table VII.

The mean apparent error rate for all regions applying Quadratic discriminating score with equal probabilities was 17.60%.

TABLE VII

QUADRATIC DISCRIMINATING SCORE WITH EQUAL PROBABILITIES CLASSIFICATION RESULTS

Region	Error I	Error II	APER
SDC	20.09%	0.00%	19.53%
SDL	11.65%	0.00%	9.75 %
SDN	21.73%	0.00%	15.87%
SDO	58.00%	0.00%	48.39%
SDT	6.90%	6.25%	6.76%

where:

- Error I: transformers from group 2 classified as group 1;
- Error II: transformers from group 1 classified as group 2;
- APER: apparent error rate.

C. Quadratic discriminating score with a priori probabilities results

As occurred with Quadratic discriminating score with equal probabilities, the application of Quadratic discriminating score with *a priori* probabilities has satisfactory results only for four regions. This problem is related to the database matrix, so West Distribution Region (SDO) data didn't reach a satisfactory result.

The best classification result was obtained for North Distribution Region (SDT) with an apparent error rate of 6.76% and the worst result was West Distribution Region (SDO) with an apparent error rate of 48.39% as shown in table VIII.

The mean apparent error rate for all regions applying Quadratic discriminating score with equal probabilities was 17.04%.

TABLE VIII

QUADRATIC DISCRIMINATING SCORE WITH *A PRIORI* PROBABILITIES CLASSIFICATION RESULTS

Region	Error I	Error II	APER
SDC	20.09%	0.00%	19.53%
SDL	10.68%	0.00%	8.59 %
SDN	19.57%	0.00%	14.29%
SDO	58.00%	0.00%	48.39%
SDT	5.18%	12.59%	6.76%

where:

- Error I: transformers from group 2 classified as group 1;
- Error II: transformers from group 1 classified as group 2;
- APER: apparent error rate.

D. Final results

Table IX shows the apparent error rate for all regions for three methods applied.

TABLE IX
APER RESULTS FOR ALL REGIONS

Region	Logistic Regression	Quadratic Score with equal Probabilities	Quadratic Score with *a priori* Probabilities
SDL	**8.40%**	9.16%	**8.40%**
SDT	18.90%	**6.75%**	**6.75%**
SDC	**6.25%**	18.75%	18.75%
SDO	**16.13%**	48.39%	48.39%
SDN	25.40%	15.87%	**14.28%**
Mean	14.64%	17.60%	**17.04%**

It can be observed that:

- Logistic regression shows the best performance for voltage THD distribution transformers classification;
- Quadratic discriminating score with a priori probabilities hasn't a better performance than Quadratic discriminating score with equal probabilities;
- The West Distribution Region (SDO) has great influence in classification performance for Quadratic discriminating score with a priori probabilities and Quadratic discriminating score with equal probabilities.

For each region it can be observed:

- SDL: East Distribution Region has a better performance with Logistic regression;
- SDT: North Distribution Region has a better performance with Quadratic discriminating score with equal probabilities;
- SDC: Center-South Distribution Region has a better performance with Logistic regression;
- SDO: West Distribution Region has a better performance with Logistic regression;
- SDN: Northwest Distribution Region has a better performance with Quadratic discriminating score with *a priori* probabilities.

VII. CONCLUSIONS

The growing demand for better energy service quality is observed by consumer's requirements for regulatory agencies. This paper proposes a statistical methodology for distribution transformer voltage THD prediction trough technical data and energy use information. The methodology is characterized by discriminating, pattern recognition and classification techniques usage. The developed models were based on distribution transformers monitored data from COPEL power system.

The methodologies were applied for transformer classification and recognition for all regions from COPEL power system (SDL, SDC, SDO, SDT, and DSN) in three different manners, using logistic regression, quadratic discriminating score with equal probabilities and quadratic discriminating score with *a priori* probabilities.

The applied models use historical information of voltage harmonics from COPEL monitoring campaign and transformer's technical characteristics. Moreover, it was considered energy use information as type and number of consumers, as well as, their energy consumption.

The results obtained proved the efficiency of the statistical model even with a small sample to its development. The classification error achieved shows the chosen characteristics influence in voltage THD predicted. This model allows evaluate all distribution transformers of COPEL power system to find the ones with PQ problems, so it will be possible plan a mitigation action.

The methods developed can be used as an efficient instrument for distribution transformers pre-selection for monitoring campaigns, this can be a good economic advantage, and since the operational costs of monitoring distributions transformers are high.

VIII. REFERENCES

[1] J. Arrilaga and N. R. Watson. *Power Systems Harmonics.* New York: Jonh Willey & Sons, 1982.

[2] G. H. Duntfman. *Introduction to Multivariate Analysis*. Beverly Hills: Sage Publications, 1984.

[3] ONS (Power System National Operator). *Procedimentos de Rede, Submódulo 2.2, Padrões de Desempenho da Rede Básica*, available at http://www.ons.org.br.

[4] ANEEL (Brazilian Electricity Regulatory Agency). *Resolução Nº 505, de 26 de novembro de 2001,* available at http://www.aneel.gov.br.

[5] R. A. Johnson and D. W. Wichern. *Applied Multivariate Statistical Analysis.* 4th ed. New Jersey: Prentice-Hall, inc., 1998.

[6] IEEE Std. 519 – 1992, - IEEE Recommended Practices and Requirements for Harmonic Control in Electrical Power Systems. 1992

[7] IEEE Std 1159-1995, IEEE Recommended Practice for Monitoring Electric Power. 1995.

[8] P. P. Barker, J. J. Burke, *et al*. *Power Quality Monitoring of a Distribution System*. IEEE Transaction on Power Delivery, Vol. 9, No 2, April 1994.

[9] M. Ungureanu, and M. Scutariu. *Power Quality Measurements on Industrial Distribution System*. 8th IEEE ICHQP, Greece. October 1998.