EVALUATION OF TECHNICAL LOSSES ESTIMATION IN LV POWER DISTRIBUTION SYSTEMS

Leonardo QUEIROZ  ANEEL – Brazil  leonardoqueiroz@aneel.gov.br

Celso CAVELLUCCI  UNICAMP – Brazil  celsocv@densis.fee.unicamp.br

Christiano LYRA  UNICAMP – Brazil  chrlyra@densis.fee.unicamp.br

ABSTRACT
This paper proposes a methodology to estimate technical losses in low voltage (LV) distribution systems. Its main contribution is the development of regression models able to estimate technical losses with low levels of information about the network. Regression models are used to predict one variable (dependent variable) from one or more variables (independent variables). The dependent variable is the technical loss, and the independent variables are any data that allows characterizing a LV network. Different sets of independent variables are evaluated, with respect to two scenarios: a distribution company and the regulatory agency. Independent variables are analyzed with respect to relevance (given by the value of statistical correlation), availability and auditability. An analysis of the regression models and of the set of independent variables is also presented.

INTRODUCTION
Since the fifties, with the increasing popularization of the digital computer, power flow models have been developed to determine the status of the systems – voltages and currents. In distribution systems it has been more common since 1980, with some approaches that take advantage of the radial configuration of distribution networks [1]. Since then, the usual approach to evaluate technical losses is based on a power flow algorithm, sometimes improved with on-line measurements at strategic points. However, there are many Brazilian distribution companies that still do not have a power flow algorithm available for studies of low voltage (LV) networks. Indeed, for some municipal distribution companies with less than 3,000 customers, it is not economically attractive to implement some computational tools. Also, some larger distribution companies have poor information about their LV networks, what turns difficulty to assess system performance.

Furthermore, the Brazilian Electricity Regulatory Agency – ANEEL needs information about technical losses for each Brazilian power distribution company; this is a key information to establish reference prices for energy. Therefore, ANEEL needs an estimation of technical losses in power distribution companies. Additionally, it is desirable to assess all distribution companies with the same methodology.

The estimation of technical losses with low level of information in LV networks has already been handled in previous papers [2-5]. In most approaches the motivation was the loss estimation for companies that do not have enough information about their networks. These papers proposed a variety of approaches, from simple equations to load flow analysis in representative networks. The present paper proposes a new approach, which relies on regression models to estimate technical losses with low level of available information.

AN OUTLINE OF LOSSES IN BRAZIL
Technical losses are part of the process of production-transportation-consumption of electrical energy. Since they cannot be eliminated, it is necessary to consider them in the planning and design of networks. It must be remarked that technical losses do not need to be reduced at any cost, but optimized to the right compromise. In a regulatory analysis, ANEEL needs to stimulate the efficiency of the electric market. Specifically in distribution systems, companies need to follow guide lines for quality and efficiency. When comparing Brazilian technical losses, we would expect they were bigger than the majority of other countries.

The first reason is the generation characteristic: the system is predominantly hydro-generation, and the hydraulic power plants are usually far away from the main consumption centres – and this scenario will get worse, since the river falls near consumption centres are almost over. So, we need big transmission lines connecting loads to generators.

The second reason for higher technical losses in Brazilian networks is the low level of load density. In addition to the large size of the country, most of the population is concentrated in the central-south regions.

Figure 1 shows the Brazilian technical and non technical percent losses of 54 distribution companies. Losses in the National Connected System – the HV transmission lines – are not computed. The mean technical losses in distribution companies seem to be in an expected level. However, Brazil has critical non technical losses levels. This is a social-economic problem, and even with some recent efforts, the problem does not seem to have a short term solution.

ANEEL adopts different regulatory treatment for technical and non technical losses. For technical losses, ANEEL applied a recent methodology to compute segmented losses, according to the voltage levels. In order to define acceptable levels for non technical losses, ANEEL uses a regression analysis with some socio-economics attributes that are somehow correlated to non technical losses, and it is establishing target levels for distribution companies.

It is important to have an adequate method to estimate technical losses, both for its own sake and because non technical losses are estimated as the difference between total losses and technical losses.
LV TECHNICAL LOSSES ESTIMATION

Availability of Information

This paper focuses on the estimation of technical losses in an environment of low level information about the network. Two scenarios are studied:

- The first scenario studies technical losses in distribution companies with poor knowledge about their LV networks (different degrees of poor knowledge are considered);
- The second scenario evaluates technical losses from a regulatory point of view; in this scenario, the information required by the methodology must be available for all distribution companies and must be auditable by ANEEL.

These two scenarios were not explicitly developed, but can be seen in the different degrees of information shown in the sets of variables required by the model. These sets were built from the following available variables for each LV circuit:

1. Nominal voltage: voltage of the circuit (V);
2. Minimum resistance: resistance of the best cable (ohms/km);
3. Maximum resistance: resistance of the worst cable (ohms/km);
4. Length: total length (metres);
5. Number of customers: number of the customers of the circuit;
6. Consumption: total consumption of the customers during the period (kWh);
7. Maximum demand: maximum coincident demand registered at the transformer (kVA).

The data set used has information from 1082 LV distribution networks [6].

Given the available variables, we can combine them in different sets — simulating a variation of the level of information. Table I shows the sets proposed — they are sorted according to the crescent level of required information. The variables are referred by the numbers assigned in last paragraph.

The classification according to relevance and availability is somehow subjective, and it can vary among companies. Variables can be classified according to their nature: network (variables 1 to 4) and load variables (5 to 7). Network variables are easier to obtain, since they are more static than load variables, which can be obtained by measurements (desired) or estimation procedures.

Variable 5 needs to be used with caution: it does not represent any load, but has high correlation with consumption and maximum demand (0.877 and 0.902 of Pearson Correlation Coefficient, respectively). But it may vary from company to company, or even for different regions of a company; its use is recommended only for companies that have a uniform LV market.

Some other variables were available for the study, but were discarded. For instance, the maximum voltage drop of each circuit is highly correlated to energy losses (0.8 of Pearson Correlation Coefficient), but it is one of the most unavailable variables and it is very difficult to audit. The nominal power of transformers (KVA) has a good availability, but it is influenced by the design criteria of each company, and its use may lead to wrong results among companies.

As Table I shows, sixteen sets of variables were considered. The first (A) represents the lowest level of information about the network — notice that no explicitly information about load is present. Explicitly, because it may carry it implicitly in design criteria, such as maximum current, that regression methods may “learn”. The last set (P) represents the highest level of information.
In linear regression analysis, transformations of variables are recommended due to linear regression requirements, such as linearity and normality. For this reason, the following transformations were attempted (for independent and dependent variables): logarithm, square root, square power and inverse. Transformed sets of variables will be represented by $t$, i.e., $A_t$ will represent $A$ set with transformed variables. Thus, we have 32 different sets of independent variables ($A$ to $P$ and $A_t$ to $P_t$), and 2 sets of dependent variables (losses and transformed losses, $Y$ and $Y_t$), composing 64 different sets of variables (including independent and dependent variables).

Independent variables were analyzed with respect to relevance for energy losses (given by the value of its statistical correlation). The correlation analysed was the Pearson Coefficient, the Spearman’s and the Kendall’s Rank. The first gives a measure of linear relationship between each variable and the energy loss, varying from -1 (perfectly opposite correlated) to 1 (perfectly correlated). The null relationship is indicated by 0 (zero). Table II shows the Pearson Correlation between each variable and energy losses (for original and transformed variables).

Since we know that there are nonlinear relationships in energy loss functions, the Kendall’s and Spearman’s coefficients were used. They are not parametric methods, and their measurement of relationship is given as a rank. Both ranks corroborated the results of Table II. The correlation coefficient is a measure of relationship. Therefore, in spite of being a good indication, it cannot be interpreted isolated. Additional analysis, as always recommended in statistics studies, need to be done. Actually, the availability and auditability were also considered. Indeed, it does not matter if a variable has a direct relation with losses if it is not available and auditable (for instance, the maximum voltage drop).

Regression Models
Once the sets of variables are defined, models of regressors were developed to make the inference between variables and energy losses. Regression models are used to predict one variable (dependent variable) from one or more variables (independent variables) [7]. The regression models investigated are:

- Linear regression (LR);
- Robust regression (RR); and
- Artificial Neural Networks (ANN).

Even though LR and RR are linear models, while LR uses the least squares method to fit the data, RR uses an iteratively reweighed least square – which is less sensitive to outliers.

The ANN used is a Multi-Layer Perceptron (MLP) [8], with the Levenberg-Marquardt optimization rule to update weights. ANN, especially of the MLP type, is widely used in regression analysis, due to its characteristic of universal function approximation.

After many simulations, the ANN configuration chosen was one hidden layer with 10 neurons, and 5000 iterations were used as maximum iterations stop criteria. To train regression models, the 1082 LV distribution networks were divided in 757 for training and 325 for test.

### Table I: Sets of the seven variables.

<table>
<thead>
<tr>
<th>Set</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table II: Pearson Correlation Coefficient.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Original</th>
<th>Transformed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.246</td>
<td>0.276</td>
</tr>
<tr>
<td>2</td>
<td>-0.326</td>
<td>-0.420</td>
</tr>
<tr>
<td>3</td>
<td>-0.013</td>
<td>-0.040</td>
</tr>
<tr>
<td>4</td>
<td>0.632</td>
<td>0.693</td>
</tr>
<tr>
<td>5</td>
<td>0.675</td>
<td>0.722</td>
</tr>
<tr>
<td>6</td>
<td>0.769</td>
<td>0.850</td>
</tr>
<tr>
<td>7</td>
<td>0.773</td>
<td>0.854</td>
</tr>
</tbody>
</table>

### REGRESSION ANALYSIS

Because of the impossibility of presenting here all the results (there are 192 results to be shown: 3 regressors executed 64 different sets of dependent and independent variables), Table III presents only the ten best regressions found, ordered from the first to the tenth. All of them were given by ANN. The Root Mean Squared Error – RMSE is the measurement of goodness of fit used. Column “TLD” – Total Loss Deviation – gives the percent difference between total loss of test set calculated by regressors and total real loss. “Set X” refers to input variables (independent), while “Set Y” refers to output variables (dependent).

As expected, ANN found the best results. But it is important to highlight that linear regression obtained good results too. Table IV shows the mean performance of the three regressors. Another assessment of performance can be presented by the data sets. Table V presents a comparison between the 32 independent variable sets.
Note that the lack of information about load, absent in variables $A$ to $D$, may explain the worse performance of these input sets. On the other hand, the maximum information about loads, present in sets $M$ to $P$, does not seem to be mandatory: only two sets that contain them ($P$ and $M$) are present in the ten best sets. Other inference is about transformations: sets with transformed variables have worse performance than sets with the original variable – although Table III has a lot of transformed sets in the best results. It is important to emphasize that Tables IV and V contain mean results, and are “polluted” with other aspects.

CONCLUSION

This paper presented a methodology to estimate technical losses in LV distribution systems. The aim is to predict energy losses with low level of information. Although the crescent uses of load flow tools, this methodology is applicable to those companies that have low level of information. They can apply directly the methodology or submit their data to a pre-processing step, grouping networks in representative clusters and applying the methodology to one network that represents each cluster. The methodology is also indicated for the regulatory agencies, which need to estimate losses of all companies using the same methodology.

Three regression methods were tested. While ANN achieved best results, LR showed that it can be used due to its simplicity. An analysis of variable sets showed that it is important to use load variables, even considering the difficulty to audit it. But while maximum coincident demand is hard to audit, the energy consumption is more auditable.

Acknowledgments

The authors would like to acknowledge the support given by the Brazilian National Research Council (CNPq).

REFERENCES


