A LINEAR MULTIVARIABLE REGRESSION MODEL TO CONSIDER THE IMPACT OF DISTRIBUTED GENERATION IN THE POWER FORECASTING: METHODOLOGY ON ENEL DISTRIBUZIONE MV ELECTRICAL NETWORK

Francisco Jose CARO MORON
Enel Distribuzione S.p.A. – Italy
Franciscojose.caromoron@enel.com

Antonio PANICO
Enel Distribuzione S.p.A. – Italy
antonio.panico3@enel.com

Domenico DI MARTINO
Enel Distribuzione S.p.A – Italy
domenico.dimartino-dd@enel.com

Marco DI CLERICO
Enel Distribuzione S.p.A – Italy
marco.diclerico@enel.com

Alessandro DE NANDO
CESI S.p.A. – Italy
alessandro.denando@cesi.it

David DRESCO
CESI S.p.A. – Italy
david.dresco@cesi.it

Simone Matteo STORI
CESI S.p.A. – Italy
simone.stori@cesi.it

ABSTRACT

In recent years the large spread of distributed generation, with a growth that is not comparable with that of pure "passive" load, has deeply changed the electrical system, even requiring new criteria for loads forecasting process.

This paper shows the forecasting process of power flow on HV/MV transformers, i.e. on the points of interconnection between the transmission and the distribution network. Furthermore, new methods and algorithms, developed and adopted by Enel Distribuzione, are here described.

In particular, starting from the energy data, measured on the MV side of HV/MV transformers, new implementations have been studied in order to separate pure "passive" load from generation, due to their increasingly independent trends, thus allowing a better estimate of energy and power flows on HV / MV transformers in primary substations.

Enel Distribuzione and CESI, in order to evaluate the pure load, have studied a procedure based on the estimation of the coefficients at national and regional level. The estimation of coefficients is carried out by means of a multi-variable linear regression model as a set of macroeconomic terms.

Furthermore, regarding to the forecasting process, by estimating the power generation that will be connected in the following years, for each energy source, it will be possible to predict potential saturation problems that may occur in primary substations in the different load / generation scenarios.

Three different scenarios have been defined: pure "passive" load, pure "passive" load / maximum generation, pure "passive" load / minimum generation, in order to evaluate, for each scenario, the potential saturation problems for the next years with monthly detail at least.

Obviously, the estimate of generation to be connected in the following years is based on available information, thus depending on legislative framework changes. In any case, by studying different scenarios, it is possible to achieve absolutely precautionary results even in case of significant changes in the legislative framework.

The methods that have been developed, described in this paper, are based on the opportunities offered by new technologies, with particular regard to energy metering systems installed in Enel Distribuzione primary substations and in the connection points for each generator, and on the availability of typical DG energy generation diagrams.

Finally, the new methods have been applied for the upgrade of "PRECAR", the software application used by Enel Distribuzione for energy flows forecast, thus allowing a higher efficiency and reliability of results.

INTRODUCTION

Load forecast is a basic tool for electrical network planning. The paper shows the basic criteria for load forecasts, the calculation algorithms adopted by Enel Distribuzione and the implementations necessary to take into account the large amount of energy produced on the MV and LV networks from renewable sources.

In fact, the spread of distributed generation (DG) connected on MV and LV networks, in terms of power connected, has become, especially in certain areas, a key factor for network planning.

Further factors complicate, then, the reference scenario:

- plurality of network operators;
- free energy market;
- continuous development of the technical standards and of legislative and regulatory framework.

In this context, the methodologies proposed and described in this paper are based on the possibilities offered by technological developments, with particular reference to energy and power measuring tools already adopted by Enel Distribuzione.

These methodologies are the starting point for the upgrade of "PRECAR", the software application used by Enel Distribuzione for load forecasts.

MEASUREMENT DATA ORIGIN

The HV electrical network is composed by a series of nodes, connected by HV electrical lines.
Each node represents an HV network point in which one or more entity (a primary substation, an HV user, a power plant, a point of energy exchange with other operators) are connected, injecting (generation) or absorbing (load) energy from the HV network.
In each node electrical measuring tools (Measurement Points, PDM) are installed.
The energy measured by a PDM may be:

- absorbed (flowing from network toward users);
- injected (flowing from users toward network).

Every PDM provides, every 15 minutes, measurements of active and reactive energy in the 4 quadrants.
Concerning the calculation of the energy flow forecasts and the possible critical issues, only the data measured by PDM installed on MV side of transformers, in Primary Substation, are involved.
The PDM are all bi-directional, so they are able to measure the total exchanged energy, between the NTN and the MV distribution network, in both directions.
This exchanged energy is the difference between the total energy required by the loads and the total energy generated on MV and LV networks.
The measured parameters are:

- injected active energy (E_c), conventionally considered with a negative sign;
- absorbed active energy (E_p), conventionally considered with a positive sign;
- injected reactive inductive energy (E_{rcL}), conventionally considered with a negative sign;
- absorbed reactive inductive energy (E_{rpL}), conventionally considered with a positive sign;
- injected reactive capacitive energy (E_{rcC}), conventionally considered with a negative sign;
- absorbed reactive capacitive energy (E_{rpC}), conventionally considered with a positive sign.

HISTORICAL DATA

The following energy values, for every 15 minutes, are stored:

- injected active energy (E_c);
- absorbed active energy (E_p);
- injected reactive energy (as a result of E_{rcL} - E_{rcC});
- absorbed reactive energy (as a result of E_{rpL} - E_{rpC}).

For each PDM, the following values are calculated, on monthly and annual basis:

- Annual Active Energy (E), expressed in MWh, that is the algebraic sum, on annual base, of E_p and E_c;
- Annual reactive Energy (E_r), expressed in MVArh, that is the algebraic sum, on annual base, of E_{rp} and E_{rc};
- Maximum active power (P_{max}) and its peak instant [MW];
- Minimum active power (P_{min}) and its peak instant [MW];
- Maximum reactive power (Q_{max}) [MVAr];
- Minimum reactive power (Q_{min}) [MVAr];
- Maximum apparent power (P_{appmax}) and its peak instant [MVA];

By the same procedure, starting from the stored values, the same parameters are calculated also for geographical entities (regions and provinces).
Yearly energy values for a certain entity (Primary Substation, province, region) constitute the historical series (E^{CP}, E^{PR}, E^{Re}) for the different entities. For each Primary Substation and for every year, the active powers, simultaneous with the annual maximum and minimum regional power, are also calculated for network planning purposes.

Separation of pure load and generation components

Starting from the $E^{CP}(t)$ values, new implementations have been defined in order to achieve the separation of pure load and generation components; in fact, because of their different nature, they have different trends and are mutually independent.
Achieving such separation is possible, for each primary substation, by analyzing the available historical data in terms of produced energy (with a granularity of 15') and in terms of power installed.
Moreover, data is also available with reference to the following different energy sources:

- photovoltaic;
- wind;
- other sources.

The separation of historical values will enable the estimate of energy and power flowing on HV/MV transformers in primary substations, with reference to the pure load.

PURE LOAD FORECAST, BASED ON HISTORICAL DATA

As said before, starting from the historical energy data, measured by PDM, and the produced energy, it's possible to obtain the pure load energy components (E_{ch_prev}).
Pure load forecast calculations are mainly based on linear interpolation of the available historical data in order to extrapolate energy and power values for the next 5 years.
In order to successful calculate the interpolation, some conditions shall be met:

- availability of complete historical data for at least 3 years;
- active energy fluctuations over the years must stay within given thresholds, apart from the case of disposals/acquisitions of network assets or putting into service new transformers;

- energy flows in a given entity cannot be constant for 2 or more consecutive years; otherwise method of interpolation is considered not applicable.

If the above conditions are not met, following method ”A” will be applied, while in the other case it will be used the method “B”.

A. Calculation methodology with historical series < 3 years

The load growth rate is estimated applying a preset annual rate.
This rate is generally set as the historical rate, previously calculated for the same province, or on the basis of load evolution foreseen in the area, for example, due to a new born industrial area.

B. Calculation methodology with historical series > 3 years

The historical data series for a certain entity (Primary Substation, Province, Region, ENEL Distribuzione network) embodies the information about energy or, for example, maximum power in the previous years. So, it’s the basis for the forecasting process.
The prediction of energy values is performed by historical data series interpolation and, consequently, by extrapolation of the future series.
The interpolation is performed using classical formulas for linear interpolation in which the coefficients are obtained by the historical series of pure load values.
The forecast of annual active energy, at different levels of aggregation, is performed as described in the following steps:

- extrapolation of the future series of annual active energy, for all levels of aggregation;
- adjustment of the results of extrapolations for each province to regional forecasts;
- adjustment of Primary Substation load forecast to provincial predictions.

Adaptations of steps 2 and 3 shall ensure that, for each year, the sum of the estimates for Primary Substation energy corresponds to the prediction of the province, and the sum of the forecasts for provinces matches the predictions of the region.

ENERGY GENERATION FORECAST, BASED ON HISTORICAL DATA

Starting from power and energy historical values for DG connected to each primary substation, it’s possible to calculate the coefficients (k_h), for the three above mentioned energy sources, as the ratio between average energy and the corresponding installed power.
These coefficients will be used to calculate, after estimating the expected DG connections, the energy production with detail on the quarter-hour, for each different source, for the next years.
The formula that is used to evaluate the total energy, generated by the different energy sources, is:

$$E_{gdh\_prev=} (P_{inst31/12/yyyy} + P_{futura}) * k_h$$

where:

- ($P_{inst31/12/yyyy}$) is the historical value of installed power;
- (P_{futura}) is the power of expected DG connections.

GLOBAL FORECAST AND CRITICAL ISSUES IN HV/MV NODES

As said before, the large increase of DG connected on MV and LV networks, in terms of power connected, must be taken into account, in order to achieve a more effective network planning.
For instance, in some cases, DG can sustain the network by means of feeding the loads that are connected at the same voltage level, thus unloading lines and transformers; in other cases DG can cause problems of saturation on HV/MV transformers.
Said that, in order to identify all the possible critical issues on HV/MV nodes, both in terms of saturation of HV/MV transformers and of reverse energy flow operations, forecasts must be evaluated in, at least, three different combinations of load/generation (scenarios):

- pure load (E_{ch_prev});
- pure load and minimum generation (sum of E_{ch_prev} and $E_{gdh\_prev} * \alpha$ where $0<\alpha<1$ is a reduction coefficient);
- load and maximum generation (sum of E_{ch_prev} and E_{gdh_prev}).

For each different scenario, it’s possible to verify:

- the reverse energy flow operation on a certain Primary Substation (the average power the quarter-hour is negative);
- overloads in a Primary Substation, that can be evaluated according to the overload capacity threshold in a “N” and “N-1” condition.

The above described methodology is applied by Enel Distribuzione using the new version of “PRECAR”, that takes into account the contribution of both DG (existing and to be connected) and pure load with its natural evolution.

EXAMPLE OF THE NEW METHODOLOGY APPLICATION

The following table shows the available data for the Primary Substation called “Lizzano” by means of the new forecasting methodology in the pure load and maximum generation scenario.

Years	Ec+Ep [MWh]	Ec [MWh]	Ep [MWh]
2010	69.465,70	-1.100,70	70.566,40
2011	62.178,90	-15.127,10	77.306,00
2012	50.663,60	-32.376,10	83.039,70
2013	50.336,91	-39.107,09	89.444,00
2014	52.244,40	-43.436,25	95.680,65
2015	54.632,60	-47.284,70	101.917,30

Table 1. Historical and forecast energy data for the primary substation “Lizzano”

In particular, the first column shows - for years 2013, 2014 and 2015 - the resulting sum of pure passive load and DG contributions, that is no more comparable to pure load forecast, due to the increasing DG contribution.
This difference is also shown in the following graph, where the qualitative behavior of the single components are separately drawn; the blue line shows, for years 2010-2012 the trend of total energy measured by PDM installed in the Primary Substation "Lizzano" and for years 2013-2015 the forecasted values.
In this scenario is evident as the effect of DG contribution reduces the power flow in the PDM and it's quickly growing from one year to another.

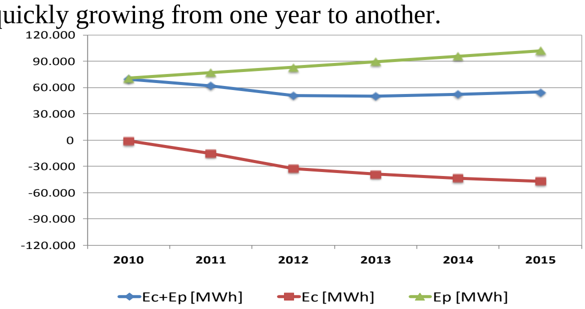

Figure 1. Energy trend in the primary substation "Lizzano"

MULTI VARIABLE LINEAR REGRESSION MODEL

Another tool that is currently used for the pure load peak forecasts is a linear interpolation model. This kind of model is quite effective to foresee the consumption data when there is a stable economic trend.
In Italy the economic trend ratio has been relatively steady (although very low) until 2008-2009. Since then the economic cycles has become more variable, with several growth and crisis phases.
Furthermore, this changing trend changes greatly between different geographic areas. This kind of economic trends are expected for the future years too.
Due to these reasons a new load forecast model has become necessary. The chosen one is a regression linear model whose inputs are the economic forecasts for the next years.
Beside of the historical load data, the new tool uses, in fact, a set of economic indicators (such as GDP growth, unemployment rate, etc.).
The linear regression tool tries to search a mathematical connection between the past load data and the set of historical economic indicators. Once defined this model, the tool uses economic forecast (provided by official institutions) to foresee the load data for next years. This procedure is applied separately to every area and to the whole Italian territory.
In the following paragraph the mathematical description of the linear regression model is provided.

Mathematical description of linear regression model

Let's define α and β_i as the Italian and regional growth coefficients of the peak loads (P_{max}) for the next year; such coefficients must satisfy:

$$P_{max,I,k+1} = P_{max,I,k}\ \alpha = \sum_{i=1}^{N} P_{max,i,k}\ \beta_i = \sum_{i=1}^{N} P_{max,i,k+1}$$

i.e. the sum of the forecasted $P_{max,i}$ at regional level at year k+1 must be equal to the forecasted $P_{max,I}$ for Italy at year k+1. Note that N represents the total number of Italian regions.
The coefficients could be obtained by estimating each $P_{max,i,k+1}$ and then:

$$\beta_i = \frac{P_{max,i,k+1}}{P_{max,i,k}} = \frac{\hat{f}(\hat{X}_{k+1})}{Y_k}, \forall i \in N$$

where the model $\hat{f}(\hat{X}_{k+1})$ is a function of a set of macroeconomic parameters $\hat{X}$ forecasted for the year k+1.
The first step of the analysis consists of creating a set of macroeconomic data to identify the model:

$$Y = \hat{f}(\hat{X}, \vartheta) = \vartheta\hat{X}$$

by means of linear least squares, and having previously selected $\hat{f}$ attributes (class, order, parameters).
In order to evaluate a large number of candidates $\hat{f}$ functions, let's consider the class of polynomial models, such as the following:

$$y = \vartheta_0 + \vartheta_1 x_1 + \vartheta_2 x_1{}^2 + \vartheta_2 x_1 x_2{}^2 + \cdots + \vartheta_j x_t{}^3 + \cdots$$

and restrict their order for computational efficiency; it is possible then to implement a procedure for combining and evaluating the regressors and choosing the best $\hat{f}$ according to some criteria.
The selected methodology is the stepwise regression, that iteratively combines ϑ_j and x_t and solve the linear regression by least squares:

$$\vartheta = (\Phi'\Phi)^{-1}\Phi'Y$$

In order to rank the models a figure of merit is needed, such as:

$$RSS = J(\vartheta) = \sum_{i=1}^{h}(y_i - \hat{y}_i)^2$$

Nevertheless, the residual sum squares (RSS) would bias the results towards the most complex model, preferring polynomials with higher grades and number of variables; for this reason, the Final Prediction Error (FPE) and Akaike Information Criterion (AIC) have also been computed to select the best function:

$$FPE = \frac{h+q}{h-q}RSS\ , \quad MDL = \frac{\ln(h)\,q}{h} + \ln(SSR)$$

The selected function (one for each Italian region and one for Italy) should then include the most influential macroeconomic variable to estimate the expected peak load; as for having to evaluate the future value of P_{max}, it is required to consider the future values of such variables as well.
The Italian National Statistics Institute (ISTAT) provides most of the future values of the variables considered during the stepwise regression; when possible, ISTAT source was used as input. Some missing data, however, have been estimated using ARIMA procedures.

Validation test case

The pure load available data are referred to the years from 2003 to 2012, for this period the following economic indicators are needed: GDP, Added Value, Population, Exports value and Unemployment rate.
Both historical and foreseen economic indicators are provided by ISTAT for the Italy, while the prediction at regional level has been calculated using ARIMA techniques on historical time series.
It's important to verify that the regression model is able to foretell correctly the annual load peaks for the next years. However, this check can be made only for the past years, whereby the load peak values are already known.
The regression model has therefore run on a shorter historical data period (2003-2010) and has foreseen the 2011-2012 two years period.
In this way is possible to compare the real and the forecasted peaks load and to assess the precision of the regression tool in comparison with that of the interpolation model. The results of the interpolation model are available only for the year 2012.
In the next graphic is possible to observe the results of the comparison, in which the error ERR% has been calculated as follows:

$$Err_{\%} = \frac{P_{\max forecast} - P_{\max real}}{P_{\max real}} \cdot 100$$

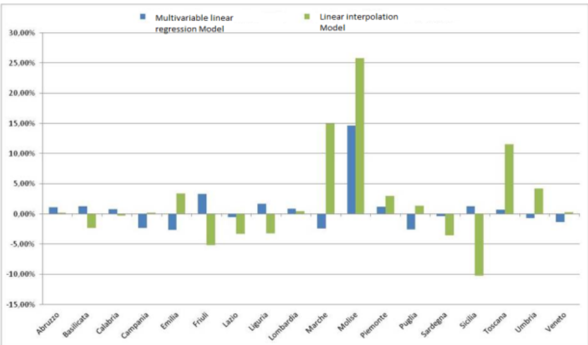

Figure 2. Error % comparison between the two models proposed

Note that the regression model and the interpolation model work on different measured values. That is due to slight differences in the method of calculation used to get the peak load values.
The prediction provided by the regression model are clearly better of those calculated with the linear interpolation. In particular for the 18 regions tested, the prediction provided by the regression model is more reliable of 12 regions (66%) than the linear interpolation model.
The use of economic factor in the procedure helps to calculate a good approximation of the pure load phenomena.
Clearly, these results are calculated with ex-post economic indicators to show methodology effectiveness, otherwise the regression model operates with foreseen economic indicators as input, which inherit an unavoidable uncertainty.
Thus, statistical estimation of the indicators is paramount for obtaining the most accurate prediction of peak loads and minimizing the overall error.

CONCLUSIONS

The large spread of distributed generation has deeply changed the reference scenario for the electrical system, with a strong impact on network planning criteria.
The opportunities offered by available technologies, with particular regard to energy metering systems, and by software applications, allow to develop new methodologies for network planning, taking into account the distributed generation.
Enel Distribuzione is developing and implementing these new methodologies, that will enable more effective results of network planning process even in a more and more complex and unstable framework, linked to current and future economic trends, forecasted for future years, which are not stable.

REFERENCES

[1] G.J. Tsekouras, E.N. Dialynas, N.D. Hatziargyriou, S. Kavatza 2006, "A non-linear multivariable regression model for midterm energy forecasting of power systems"

[2] Fabio Cazzato, Simone Botton, Federico Marmeggi, 2011, "Benefits Of The Modern Energy Metering Systems For The Distribution Network Planning And Development Processes With The Diffusion Of Distributed Generation" (*CIRED 21st International Conference on Electricity Distribution*).