Abstract – This paper is about the subject of mid-term load forecasting namely the forecasting three years ahead. The methods utilized in this analysis were the regression models and spatial load approach. The first analysis is done in the spatial load forecasting environment with the help of territorial planning. As the data required isn’t still available it was carried out a second approach based on econometric models. In this second approach a relationship between the electrical consumption and social-economic variables was researched for all the Portugal townships. The best approach allowed an average error of 2,13% and an maximum error of 12,6 % for a temporal horizon of one year and in the case of three years an average error of 2,13 % and maximum error of 25,93% . Several methods of peak load forecasting are also presented and compared.

Index terms – Forecasting, Load, Spatial Load, Regression, Peak Load.

I. INTRODUCTION

The electrical load forecast plays an extremely important role in the electrical network stability. The load forecasting problem in terms of network planning aims to know beforehand the load patterns, to follow the electrical growth as efficiently as possible. The necessary improvements in the network have to be planned years ahead to ensure that the new consumption characteristics are met. This kind of advanced planning may lead to different decisions in the energy distribution lines, to the necessity of exchange some equipment and to the prevention of some problems. Although the network planning has been studied for decades the new telemetric systems and the microgeneration phenomena are giving new means and new problems to the load forecasting.

In a the first analysis it was studied the spatial load forecasting method, although forecasting the location of the new clients throw geographic planning in simultaneous with studies of the load curves evolution is an effective method, it was not used because it requires a big quantity of geographic data which are not appropriately collected in Portugal.

Another approach was used based on a multiple linear regression for a mid to long term forecast. As the construction time of a substation is in general 3 years, this time was defined as the time horizon for the forecast.

Regarding the variables used, they were collected in the INE (Instituto Nacional de Estatística) website and have a socioeconomic character. As the aim of this analysis is to find where the load is, it was used the same consumers classes and townships divisions available in the DGE (Direcção Geral de Geologia e Energia) website. There are several classes of consumers so it was made a decision to analyze only one of them, the domestic class of consumers. Different variables were tested and different analyses are presented in this paper, all of them to all the Portugal’s townships. After the estimation of the annual energy consumption it becomes necessary to analyze which is the maximum annual power in order to make a good network planning. Different approaches are presented, one of them already considers the usage of smart meters.

Regarding the text organization, in chapter II the spatial load analysis is described with detail. This chapter also includes the analysis for very different territorial planning scenarios, the distinction between spatial and temporal analyses and the measurement of errors both in quantity and in location. It is also described the main advantages of this model and the reasons why it is not used in the Portuguese case.

In chapter III it is described the load forecasting method based on multiple linear regression. It is presented two types of regression: one based on the least square method and other in a robust regression method which is adapted to cases where a large set of data is a reality, the LTS (least trimmed squares) method. Several analyses are described for the least squares case, each one of them using different variables. A simulation with the LTS method is also presented for variables in the best LS case. Finally conclusions are draw from the obtained results and the problems inherent of each type of approach are discussed.

In chapter IV are described several methods of maximum power calculation including methods which use time series with the influence of meteorological variables, methods which use load factors and probabilistic methods. It was given a special emphasis to the method which is utilized in Finland as it incorporates in its studies the utilization of measurements made trough smart meters.

In chapter V conclusions regarding this thesis are presented and future approaches are suggested to handle the load forecasting for distribution networks planning problem.
II. SPATIAL LOAD FORECASTING METHOD BASED ON TERRITORIAL GROWTH

The main goal of the load forecasting is to find where, when and how will the energy arrive at the electrical consumers. In the case of this paper it will be analyzed the mid term load forecasting for a time horizon of three years as this is in average the construction time of a electrical substation. The spatial load forecasting must deal with two different aspects: the temporal analysis and the spatial analysis.

In the spatial analysis the goal is to identify where each type of load is located and which will be the evolution of the load in spatial terms. The electrical consumers are not the same, and react is different ways to different factors, the consumers can be divided in different commercial classes according to the power contracted and the actual power consumed. After the classes have been defined it is necessary to analyze the territory and define exactly where each type of consumers is located in spatial terms. It is essential that the data regarding the territory is in total agreement with the data regarding the electrical consumption so that it can be attributed an average consumption characteristic to each area. Another aspect to take in consideration is the division of the territory method. There are two main methods for the territorial division: to divide the map in a grip or divided in polygons. The most popular method is the utilization of the mid division, normally the data collection is done only after the division of the territory has taken place. As the polygon method, it is based on the equipment used and the system configuration determines where the areas begin and end. Both of these division methods require the utilization of the geographical information systems to determine which is the land use assigned to each area. Knowing the geographical location of each type of land it is necessary to do the correspondence between the land and the electrical consumers. After drawing a map of the actual location of all the electrical consumers it is necessary to consider how the spatial location of the consumers will change in the temporal horizon of the forecast.

Meaning that it is intended to produce a new map of the electrical consumption taken in consideration that the region will change as new investments and ventures will be made in the studied area. In the new map it should appear new urbanizations, new commercial areas, new industries and changes in the land use. After drawing the new load focus in the new map it is usually not enough to explain the load growth forecasted for the city. The H. Lee Willis [1] method sustains that the remain load should be attributed to the vacant areas. To define which empty areas will be occupied it is attributed a probability development scale t the inter map. After the attribution of an numerical scale according to the more or less influent factors, it is possible to do a map with the most attractive areas to the territorial development, such as the map presented in figure 1.

![Figure 1 – Regional development scaled map [2] .](image)

After the map is done it is possible to assign the new loads to the areas with the bigger development probability. But this close relation between economic growth and load growth can have negative effects in planning of the energy distribution network. For example, the fact of the construction of a factory ochre in one side of a city or another will affect the whole distribution network, implying totally different visions in terms of the distribution plan. The best approach to follow when it is not exactly kown where an investment will take place is to construct different scenarios in long term analyzing the impact of each one off them in the electric network.

In the temporal analysis the goal is to forecast how the load curves will change and do it for all the different consumer classes. This kind of analysis is possible when the information regarding each type for cure is available. Knowing the typical curve and how it has been it’s variation it is possible to do forecasts of how will each curve evolve in time according to the main factors that influence the load.

Regarding the measurement of errors, the precision of the forecast has to be measured in terms of a spatial error and a quantiative error. As the errors relatives to the quantity of power forecasted over or under the real value, there are two main indicators of such errors: the medium absolute error (MAE) and the mean square error (MSE). In the error regarding the location the goal is to take in account if the spot where the load is located was missed and not the quantity of load. The main idea of this kind of evaluation is to construct a map with the obtained quantity errors. To do so, historical data is consulted and two different algorithms are applied, method A and method B, then a simulation with this data takes place. After the simulation the calculus of the error is made subtracting the obtained results to the real results. The difference between those values forecasted and real is considered the region error map. This kind of error measurement will translate how far was the method of predicting the location of the load and not the quantity of load. In order to compare the different algorithms of forecast it can be measured the distance that the different loads had to travel to compensate the planning error.
Considering the influence of the spatial resolution utilized, it can be said that the smaller the area analyzed the bigger amount of data will be needed and bigger are the chances of spatial error. The error measured through quantitative methods like the average absolute error (AVE) or the medium square error (MSE) will increase as the map scale is reduced. The H. Lee Willis [1] method suggests the following golden rule: the areas with a tenth of the size of the service area are generally enough to obtain good results with the analysis of the spatial error.

Although this possible a more accurate forecast with this method, as it performs a combination between spatial forecast and temporal forecast to obtain a new consumption map, this method can’t be used in the Portuguese case. This happens because there aren’t geographical data available for such small resolutions.

### III. MODEL OF ENERGY CONSUMPTION FORECAST BASED ON LINEAR REGRESSIONS

The methodology used in this paper is the econometric model based on multiple linear regression. In this chapter it will be addressed two distinct methods of forecasting the energy consumption. The least squares approach and the robust regression approach.

In the multiple linear regression model it is described a relation between a set of quantitative independent variables $X_i$ and a dependent quantitative variable $Y$ [3]. The objective is to determine the set of parameters $\beta_i$ that link the systems answer $Y$, in this case the energy consumption, with the regression variables $X_i$. The linear equation which translates the model is the equation 1.

$$Y = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_k x_{ki} + \varepsilon_i, \ i = 1,..,N \quad (1)$$

In this formula $k$ is the number of quantitative independent variables, $i$ the number of observations, the parameters $\beta_i$ are know as regression coefficients and they translate the dependence between the answer of the system and the respective regression variable $x_i$ when all the other variables remain constant. The letter $\varepsilon_i$ represents a random error associated with the observed value of $y_i$.

#### III.1. THE LEAST SQUARE APPROACH

In chapter II, it has been tried to make an approach at local level evolving in to a global vision. This new method aims to do the opposite. The method begins from a more global forecast, which is associated with a lower error to arrive at a more local level, the substation.

In mid and long term load forecasts the system response is the annual energy consumption. The other regression variables are selected according to which type of activity sector is been analyzed. In the case of this paper the variables were chosen according to the municipal available data in the INE website.

Several distinct analyses have been performed in order to try to find out which of the socio-economic variables influence more the energetic consumption. For the cases where variables regarding distinct years have been used, these variables where treated like independent variables, such as it is described in equation 2.

$$\ln(Kwh) = \ln(A_0) + \beta_1 \ln(x) + \beta_2 \ln(y) + \beta_3 \ln(z) \quad (2)$$

In INE website there are different statistical variables at municipal level, to decide which ones should be used to calculate the energy consumption of the domestic class of consumers it was utilized the Pearson correlation coefficient $R$, described in equation 3.

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}} \quad (3)$$

This coefficient takes values between -1 and 1, being a module valor superior to 0,7 considered a strong correlation, the analyzed variables which presented a coefficient superior to 0,7 were used and all the others were discarded.

In a first analysis, as there is a bigger amount of data for the years 2004 and 2005 it were utilized these years for the variables research. It was created an algorithm in Matlab which utilizes the 10 variables with the bigger $R$ for these two years, which was able to find out which was the better combination of variables that provided a better result. For the calculus of all the different Portugal’s townships it has been verified that using only statistical data of 2005 to calculate the electrical data of 2005 obtain an medium error of 8,91% and a maximum error of 42,33%.

For the calculus of the electrical data of 2004 utilizing only statistical data of 2004 it is obtained an medium error of 4,58% and a maximum error of 24,32% Both cases for a $Y$ dependent of 7 variables. As these errors are quite high it were made the calculations of the electrical data of 2005 from the statistical data of 2004, it was obtained an average error of 2,33% and a maximum error of 13,99% for a $Y$ dependent of 4 variables. It was also calculated the electrical data of 2005 from the statistical data of 2004 and 2005 in simultaneous, it was obtained an medium error of 2,26% and a maximum error of 13,71%, for a $Y$ dependent of 6 variables. It can be verified that the best results occurs when the statistical data of 2004 and 2005 are utilized to calculate the electrical data of 2005 which shows a inertia in changing
the consumption habits from the consumers when the economic variables change as it was expected.

It was effectuated a second analysis, as the obtained results still presented an elevated error, in this analysis it were added to the variables of the analysis 1 all the global variables which had a correlation coefficient superior to 0,7. It was analyzed a total of 58 variables. The best obtained error with this variables was for a Y dependent of 5 variables with a maximum error of 13,82% and a medium error of 2,168%. In this analysis it had to be withdrawn some townships because there wasn’t enough information to make the analysis. So from 284 townships only 259 were analyzed. The calculations for the first analysis were remade so that both analyses could be compared, the results are presented in table 1.

Table 1 – Best results for the different analysis for the 259 townships.

<table>
<thead>
<tr>
<th>Analysis 1</th>
<th>Analysis 2</th>
<th>Analysis 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>13,05</td>
<td>2,22</td>
<td>12,18</td>
</tr>
</tbody>
</table>

Looking at the table it appears that there was a slight improvement in the average error and in the maximum error as well. However it is worth noting that this method is very heavy in computing terms, once simulations with an output dependent of a large number of variables takes much longer than the initial analysis.

So far it was only made an analysis on economic data and as the objective is to provide the forecast of consumption 3 years ahead, it was added to the variables analyzed the domestical electrical consumption in 2002, 2001, 2000 and 1999. Once in this analysis it will be taken into account economic variables for the year to be forecasted and the previous year and the electrical consumption registred over 4 years. The best results were obtained for a Y dependent of 8 variables.

Looking at the table 1 it may be concluded that it is more useful to add the electrical consumption as independent variables than to consider more statistical variables. A computer analysis conducted as the one done for the 58 variables becomes too demanding in terms of virtual memory for the computer and thus is slower, while an analysis as the one done for the 24 variables in fastest. So as it is simpler and produces better results it can be concluded that the analysis 3 is the best of the three cases examined.

As one of the objectives of this paper is to calculate the energy consumption in a time horizon of 3 years, and since the years with more statistics data are 2004 and 2005 but there are no electrical data for 2007 and 2008 it must be used other years. Therefore it was used the following years with a greater number of statistical data which are the years 2002 and 2003 and it was made the forecast of the consumption for the year 2006.

Thus it will be used statistical data from 2002 and 2003 and electricity consumption data from 2003, 2002, 2001 and 2000. The best result obtained by this method was an maximum error of 25.93% with an average error of 3.42% using 7 of the variables, including 3 variables related to electrical consumption. This result is reasonable given that the prediction is being made 3 years in advance, however if there were more data on economic variables, such as the proportion of purchasing power and average number of inhabitants, before 2003 it would be easier to make a forecast because these economic variables have a high correlation with the energy consumption.

### III.2. ROBUST REGRESSION METHOD

The biggest disadvantage of the least squares method is its sensitivity to outliers. The objectives of more robust estimation methods are seeking efficient estimators under a certain model so that small changes in the distribution of the sample produce small changes in estimates.

The robust method used in this study was the least trimmed squares $[4]$ e $[5]$, LTS. The LTS estimator has two fundamental characteristics: it has a mechanism for rejection of outliers and can be implemented through an algorithm of fast convergence.

The objective of this method is to reduce the error of the model, by minimizing the square of the residues that is expressed through the equation 4., where $R_i$ represents the residues sorted and each one is given through the equation 5, where $h$<N meaning less than the number of cases examined.

\[
Q = \sum_{i=1}^{h} R_i^2 \quad (4)
\]

\[
R_i = Y_i - \sum_{j=1}^{a} x_{ij} B_{ij} \quad (5)
\]

The algorithm of this method has be aim to create subsets of data, starting with initial data and repeating the calculation of the least square method until their convergence, in order to eliminate data with interference as can be seen in Figure 2. This algorithm is based on the fact that the residues of the group of data more distant from the average are much larger than the others in the group.
The key to this algorithm is the choice of values to analyze. There are two ways to select the data: choose a group at random, or to calculate the first coefficients $\beta$ and verify that these define a single plane (that is if none of the coefficients is zero). If they do not define a plane it is necessary to extend the number of observations.

After the choice of data have been made, the basic idea according to [4] is making the calculation, according to the least squares method, of $\beta$ coefficients then calculate the residues through the equation 5. Then put the residues in ascending order, equation 6 and on the basis of this order effectuate a permutation, equation 7, of the matrix $H$ calculated during the method of least squares and given by the equations 8 and 9. That is the lines of the new matrix $H$ will be sorted according to the order laid down by the residues. Then it is recalculated the method of least squares and the cycle ends only when the square sum of residues, equation 4, is equal in two consecutive iterations.

$$|r(\pi(1))| \leq |r(\pi(2))| \leq \cdots \leq |r(\pi(n))|$$ (6)

$$H_{new} := \{\pi(1), \pi(2), \ldots, \pi(n)\}$$ (7)

$$\hat{Y} = \hat{\beta} \times X = H \times Y$$ (8)

$$H = X \left( X^T X \right)^{-1} X^T$$ (9)

Although this method is better in its approach to outliners, it is more cumbersome in the number of iterations to perform. That is, for data sets of large dimensions would be necessary to analyze a relatively large number of subsets the method would become heavier in computing terms.

However there is an approach that does not fully calculate all possible hypotheses of subsets called FAST-LTS. This approach argues that if the number of observations is less than 600 the steps to be undertaken are:

- Repeat $z$ times (where $z$ is a number defined by the user):
  - Build, using the verification that the subset defines a plan, the matrix $H$.
  - Making the calculation of the residues, the calculation of $Q$, the permutation of the matrix $H$ and the calculation of the new coefficients $\beta$ twice.

  - For the 10 best values of $Q$ compute the data until there is convergence.
  - The solution will be given by the values of coefficients $\beta$ corresponding to the case with lower $Q$.

Are still tests to be made to the equation resultant of the system to verify that the model is valid. In [4], are offered the analysis of 5 types of graphics to test the resultant equation of the system. The basic premises of a linear regression model are: linearity of the phenomenon measured, the constant variance in terms of error (homocedasticity), normality of errors, independent errors, non-colinearity and absence of aberrant observations.

This model was used and all tests were made to the model were overcome. It appears that the errors in the case of least squares and the least trimmed squares have the same order of magnitude and can be therefore be used any of them. Since the method of least squares is faster and simpler to perform for the number of items in question, the choice would fall on the method 1, analysis 3.

III.3. PROBLEMS INERENT TO THE UTILIZATION OF THESE MODELS

Although the forecasting horizon is set from 1 to 3 years because it is the time of construction of a substation, due to the scarcity of data the most examined case was the forecast for a year ahead.

As the statistical data are more plentiful for the years 2004 and 2005 the estimates were concentrated in these years. The years for which the forecasts made are mostly the years of the used statistical data, then to the value of error obtained must be added the value of the error of the forecast of theses variables. As the statistical variables are real and can be collected at any time, annually or monthly it can be set trends of their development. This case was not the case for the calculation of electricity consumption of 2006 were it were used only real variables already collected in the years 2003 and 2002.

There is also another issue relative to the two methods, although it is helpful to know the electricity consumption by township it must be taken into account where the substations are located. There are cases where a substation feeds a county, or several, or even cases where several substations supply the same regions. So it is necessary an analysis by substation territorial adjustments must be made.
So the calculations of the maximum power in summer and in winter are given by the equations 12 and 13.

\[
\text{Maximum power (Summer)} = \sum_{\text{end-use}} \text{medium annual energy} \times \text{summer load factor}
\]

\[
\text{Maximum power (Winter)} = \sum_{\text{end-use}} \text{medium annual energy} \times \text{winter load factor}
\]

**IV.2. Forecast through factors that relate the maximum power with the annual energy consumption**

In Finland [7], the traditional method of estimating the maximum power is done through the annual energy consumption by Velander formula, equation 14. In this equation \( W_a \) represents the annual energy, \( k_1 \) and \( k_2 \) are the factors studied from data collected about the load curves. This formula has proved quite accurate in the medium voltage network where the number of consumers is big, but it is very inaccurate in the case of low voltage network where the number of users is lower.

\[
P_{\text{max}} = k_1 W_a + k_2 \sqrt{W_a} \quad \text{[MW]}
\]

It was initiated a study on the load, as the number of electrical customers is extremely high and the only information available is the abundant electric bill, this study involved a sampling of customers so that classes of customers could be defined. Data for the study of the load were collected through meters installed in pattern consumers, and the load curves were measured through portable meters with memory or through remote reading. The load research produced simpler models of load curves for use in applications where only energy consumption data and customers classes information is available.

However a study was also conducted in order to incorporate the actual measurements made in the electricity networks in the model. In DLE models (distribution load estimation) actual measurements are used to improve the models of classes of consumers, the goal is then to minimize the estimation errors, errors caused either by the forecast of annual energy consumption or by the made measurements. To do so it is used the method of weighted least squares where the weights are the inverse of the variances of the models and measurements.

The output of this method is a selection of load curves for different consumers and for different categories of consumers, although the class of consumers has a typical load curve the online measurements enable the processing of each customer as an individual unit. In Figure 3 is represented a diagram of a DLE model.
Figure 3 – Distribution load estimation schematics [7].

One of the main advantages of this method is that it is possible the ongoing research of load where the needs in terms if recording data are smaller compared with other conventional methods. In addition, knowledge of the specific charges gives an indication of the need for the load control and where problems could occur in the network. The DLE has a great advantage of maximum use of network capacity and it finds the most profitable target for investment and services.

The biggest disadvantage is having to make the processing of information online, which requires a great capacity for processing data and coordination between the several utilities. There is a possibility of the storage of information to be handled later but that does not provide the information to the market about the current load.

IV.3. Power forecast through statistical methods

In some methods customers are often treated as a customer type which makes them just a number and makes the individual treatment impossible. This treatment would be necessary to study smaller areas sometimes also called load pockets. The characteristics of these areas are often different from the average feature and normally there is no information about them.

To address the problem of forecasting what happens in small areas and empty areas Feinberg et al. [8] developed a model that calculates the parameters of the system and the ones of load pockets. The multiplicative models, developed by Feinberg take into account time factors such as day of the week, time of day, but also take into consideration climatic factors such as temperature and humidity.

\[ L(t) = F(d(t), h(t)). f(w(t)) + R(t) \]  (15)

The input of data that the model requires are the weather data and the hourly electric consumption. The multiplicative model resultant of this study is represented in equation 15, in this equation \(L(t)\) represents the load in the actual time \(t\), \(d(t)\) is the day of the week, \(h(t)\) the hour of the day, \(F(d(t),h(t))\) is the daily/hourly component, \(w(t)\) are the climatic data that include the temperature and humidity, \(f(w)\) is a climatic factor and \(R(t)\) represents a random error. In reality \(w(t)\) is a vector with the current and historical data, so it is taken into account that the electricity consumption does not depend only on the weather conditions of that day but also on that happened of the previous days, this happens for instance in the waves of heat that lasts for days in a row.

Feinberg [9] developed a software that uses the method described above to learn the parameters of the model and to make predictions for the next year based on historical patterns of the load. The software allows the estimation of trends and peak loads for different locations (load pockets and the complete system), modeled data, load curves, forecasts for power for the desired day and the calculation of standardization climatic factors.

IV.4. Power forecast models which incorporate the influence of temperature and of seasonalitys

The electric load analysis shows that the demand presents cycles daily, weekly and annually reflecting the economic influence, the human activity, effects of schedule and of weather. These are the mains factors to consider in order to build an efficient forecasting model.

IV.4.1. Seasonality treatment with the influence of temperature

A feature of electric load is that it displays various levels of seasonality: daily, weekly and annual. This seasonal part is mostly soft with the exception of the daylight saving changes that introduce singularities 2 times per year.

To take into account the daily and weekly seasonality the most common approach is to decompose the data by day type having each type its one characteristic pattern. In models MTF (medium time forecast), the annual seasonality can be treated by different forecast models, and it could for instance the treated by a time series in Winter and another in Summer.

This type of model is used for seasonal variation in the EDF (électricité du France), is based on the decomposition of the load \(P_i\), where \(i\) represents the indices of the observations, in two components, as it can be seen in the equation 16:

- \(P_{hi}\) is the load part which is independent of the weather, it contains the seasonalties, tendencies and the Schedule effects.
- \(P_{ci}\) which represents the part of the load that is dependent of the weather variables.
- \(\varepsilon\) is the error of the model.

\[ P_i = P_{hi} + P_{ci} + \varepsilon_i \]  [MW]  (16)
The temperature sensitive part \( f \) of the load \( P_{ci} \) can be approximated by a nonlinear model to reflect a linear increase when:

- The temperature drops to values lower than a certain value (heat gradient).
- The temperature rises to values above a certain value (cooling gradient).

The part of the load sensitive to the temperature is them composed by a part connected to the heating and another related to the cooling \([10],[11]\), and the analyzed temperature is smoothed to assure that the modulation of the temperature variations inertia is done when it concerns the temperature within the buildings.

The part of the load related to the heat takes the form shown in equation 17, where \( g_{h,s} \) is the rate of warming for the weather station \( s \) and the hour \( h \), \( r \) is the trend of the gradient, \( t_s \) is the temperature of threshold of the heatinh for the hour \( h \) and \( \sigma \) is the dispersion of heat. The load depends on the temperature but for the cooling has a similar treatment.

\[
P_{ci} = \sum_{s \in S} g_{h,s} (1 + r \cdot y_i) \psi(u_{i,s}, t_h, \sigma) \quad [MW] \tag{17}
\]

In this model are also taken into account that the different meteorological stations have different values of temperature and that the consumption related to the temperature depends not only on the temperature but also of the thermal insulation of the buildings.

\[
P_{h,j} = \Pi_{h,k} \cdot S_i \cdot R_i \quad [MW] \tag{18}
\]

As for the independent component of the climate \( P_{h,s} \) is composed of three components in a multiplicative form, as can be seen in the equation 18:

- A part of the load that incorporates the daily and weekly seasonality, \( \Pi_{h,s} \).
- A part with the annual seasonality \( S_i \).
- And a tendency \( R_i = 1 + r \cdot y_i \).

And the seasonality \( S_i \) is described for each hour by the system of equation 19, where the variables \( q_{b,p} \) are dummy variables, per hour and per period drawn to deal with the singularities introduced by holidays and daylight savings. For each hour \( h \), a certain amount of load is added or redraw depending on the analyzed period. The dates of these singularities are forecasted with a certain time gap. Still in this system \( c_{n,k} \) and \( d_{n,m} \) are the Fourier coefficients associated with the frequency \( n \) for the type of day \( k \) and \( a_{h,m} \) are coefficients for the Fourier series for the hour \( h \).

\[
S_i = G_i (q_{h,p} + F_i)
\]

\[
G_i = 1 + \sum_{n=1}^{N} c_n \cos(2\pi nj) + d_n \sin(2\pi nj) \tag{19}
\]

\[
F_i = \sum_{m=1}^{4} a_{h,m} \cos(2\pi mj) + b_{h,m} \sin(2\pi mj)
\]

This model as proven in EDF to be the best approach either in economy or in accuracy, but it requires a very careful choice of the type of the day. This model was integrated with the current used model and it presented a good performance: in the summer it showed a gain of 25% in RMSE.

### IV.4.2. Treatment of the humidity influence

In countries with a wide range of humidity as Cyprus \([12]\), it is necessary and essential to include the contribution of the relative humidity in the consumption of electricity. High levels of humidity during the summer months or the winter tend to increase the need for cooling or heating that would be needed in normal climatic conditions.

The electrical load referent to the cooling depends of the specific enthalpy \( h \) of the environment, which through psychometric graphics, can be related to the environment temperature \( T_0 \) and with the relative humidity \( \varphi \).

This kind of models were applied in Cyprus where the levels of relative humidity are sometimes quite high and it was verified that the load forecasting presented better results than the one done with the forecasting model that only contains the influence of the temperature.

### IV.5. Forecasting models that use probabilistic methods

There is still a completely different approach to the analysis of peak power in the electricity distribution networks. This approach \([13]\) calls for simultaneous analysis of the peak of a set of customers for the low voltage networks. Meaning that this method seeks to examine how and how many customers of each type will be consuming electricity at the same time.

In addressing this issue there are essentially two possible reviews: a regulatory representation, with fixed rates; and a representation where the probability of the peak load of each customer is a random variable.

In the respective 2005 regulation in Portugal was established a coefficient of simultaneity that relates the peak of a set of customers with the sum of peaks of each customer of the set. This coefficient is the product of two factors: the first factor is in a table and indicates the
simultaneity between customers connected to a node of
the network, the second factor corresponds to the
simultaneity between loads agglomerated in the nodes
and depends on the number of nodes in that network. So
in this method the peak of a set of customers is the sum
of the peaks of those customers affected by the
coefficient of simultaneity.

One problem with this method of forecasting is
that the factor is not schedule individually and does not
even suffer amendments to a number of clients over 50.
The ratio should decrease monotonically with the
increasing number of customers, which is not the case
since the first coefficient is given in steps. Beyond this
there is another problem, the fact that it is not considered
the hired power that characterizes each customer because
in each factor is considered only the number of
customers, not their power.

Although this method presents a relatively
simple complexity, it is not a good approximation of
what happens in reality, so it became necessary to
construct a method that better represented the load.

Concerning the probabilistic representation, the
characterization of a low voltage client corresponds to a
distribution where the random variable is the power
requested by the customer in the peak hours to the
network. In [13] have been studied various probability
distributions that could be attributed to the first
coefficient of simultaneity, including the beta
distribution, including the beta distribution, processing
based on Fourier transforms, but the one with best results
was the distribution of Bernoulli. In this study were
implemented two coefficients of simultaneity, but the
power on which they were applied is the peak power,
which is the contracted power affected by a factor
dependent on the one contracted power.

Comparing the two methods mentioned above,
the both methods require much information on the low
voltage customers data because it is necessary to know
the number of clients per node, per contracted power and
all the parameters of the branches (type of material,
section and length). This type of information is typically
available to the entity responsible for the distribution of
energy but requires the analysis at a small scale which is
complicated in a considerably large area. However with
the factor of simultaneity that takes into account the
power contracted in each node it can be performed a
relatively accurate analysis of the maximum power
requested to each distribution transformer.

IV.6. Adaptation of the methods to the
Portuguese case

The all methods analyzed to calculate the
maximum power have advantages and disadvantages.
Clearly the Finnish method is the one that is more
accurate as it deals with an enormous range of
information and is able to make a model for each
substation. However this method can not yet be applied
in Portugal as the smart meters are not yet installed.

As for the method three and four are also not yet
feasible since apart from taking into account sazonalities
they would also have to bear in mid that substation feeds
different sets of load due to the Portuguese network
characteristics, so they would not be appropriate.

So only the first and fifth method could be
currently adapted to the Portuguese case and since they
are independent, they can be both made for security
analysis and admitted that the value of maximum power
is the greatest value of the two.

V. CONCLUSIONS

In this work the first method discussed was the
planning of spatial load that consists in the analysis of the
territory planning to forecast where the new consumers
will be located and where will be moved the different
types of consumers that already exist. Although this
method is one of those that shows better practical
outcomes and begins to be widely used, this method was
discarded because there isn’t currently in Portugal a data
collection system for analysis of land for such small
spatial resolutions. The data is collected mostly at
municipal or regional, both territorial and low voltage
consumer data.

It was felt that now this method could not
produce satisfactory results, which only would become
more viable when more in-depth studies are conducted
about the load diagrams, studies possible with the
introduction of telemetry systems.

The next step was to follow a fairly common
approach in the load forecasting regarding the planning of
the electric networks: firstly calculate an estimate of
annual energy request followed by a calculation of the
annual peak power.

Based on statistics available on the INE website
different tests were made to determine the variables that
affect more the electricity consumption. Two different
methods were used to perform the regression and the
results between them were compared.

Although these methods have acceptable results
for the timeframe in question they have the problem that
the data and the results are displayed by township. So it
would be necessary a transition between townships and
locations of substations to achieve more precisely results
and to determine if the electrical system remains stable.

After this analysis it was discussed several
methods for forecasting power. Of the main analyzed
methods it stands out the one that utilizes load factors as
it can be used in the Portuguese case. It is also need to
highlight the Finnish method that uses factors to relate the
power consumption with the annual maximum power.
This method was based in data extracted from intelligent
systems for the energy measurement.

As future approaches, the measurement of
spatial load is without a doubt the most accurate once it
incorporates measurements made in the electricity
network to determine the evolution of the different types
of load curves. It also uses data from the geographical
planning systems to determine the geographic location of

Page 9 of 10
new consumer and the displacement of the old ones. This method can be used when the geographical information systems that already exist today, can gather more information so that they can accurately characterize the domestic consumers.

This system integrated with the method of forecasting the maximum power presented to the Finnish case appears to be the most appropriate approach to use, since it allows the use of data for the territorial planning and online data from the intelligent tele-counting to examine the load growth.

VI. REFERENCES


[7] “Load research and load estimation in electricity distribution”, Anssi Seppälä, dissertation for the degree of doctor technology in Helsinki University of technology (Espoo, Finland)


