Instituto de Engenharia de Sistemas e Computadores de Coimbra

Institute of Systems Engineering and Computers

INESC - Coimbra

João Lourenço and Paulo Santos

Short Term Load Forecasting Using Gaussian Process Models

No5 2010

ISSN: 1645-2631

Instituto de Engenharia de Sistemas e Computadores de Coimbra

INESC - Coimbra

Rua Antero de Quental, 199; 3000-033 Coimbra; Portugal

www.inescc.pt

Short Term Load Forecasting Using Gaussian Process Models

Lourenço J.M., Santos P.J.

Lourenço J. M., is with Department of Systems and Informatics, Polytechnic Institute of Setúbal IPS Rua do Vale de Chaves Estefanilha 2914-508 Setúbal Portugal, and Research Institute INESC-ID (email: jlourenc@est.ips.pt).
Santos P. J., is with Department of Electrical Engineering, Polytechnic Institute of Setúbal IPS Rua do Vale de Chaves Estefanilha 2914-508 Setúbal Portugal, and Research Institute INESCCoimbra (email: psantos@est.ips.pt).

June 8, 2010

***Abstract*— The electrical deregulated market increases the need for short-term load forecast algorithms in order to assist electrical utilities in activities such as planning, operating and controlling electric energy systems. Methodologies based on regression methods have been widely used with satisfactory results. However, this type of approach has some shortcomings. This paper proposes a short-term load forecast methodology applied to distribution systems, based on Gaussian Process models. This methodology establishes an interesting and valuable approach to short-term forecasting applied to the electrical sector. The results obtained are in accordance with the best values of expected errors for these types of methodologies. A careful study of the input variables (regressors) was made, from the point of view of contiguous values, in order to include the strictly necessary instances of endogenous variables. Regressors representing the trend of consumption, at homologous time intervals in the past, were also included in the input vector. The proposed approach was tested on real-load from three medium-sized supply electrical distribution substations located in the center of Portugal. To test the performance of the model in different load situations, the case study includes three different electrical distribution substations representative of typical load consuming patterns, namely the residential, the non-residential and the service sector.**

***Index Terms*— Gaussian Process Model, Load Forecasting, Distribution Systems, Electrical Distribution Substations, Deregulated Market.**

I.INTRODUCTION

The management of electricity distribution networks is becoming progressively more demanding. In fact, liberalization tends to eliminate some of the certitudes that formerly were taken for granted in the utility business. In a deregulated environment, commercial transactions take place with a reasonable independence of the technical issues of network management. The system operators (SOs) have to rely as much as possible on reliable data, namely on load forecast results, keeping in mind that uncertainty is a key issue in most decisions. To provide some decisive support, the SOs have to use reliable forecast results [1], [2].

Short-Term Load Forecasting (STLF) methodologies have evolved, during the last four decades [3], starting with approaches based on time series analysis through statistical methods [4].

Later, they developed into knowledge based systems, and more recently, other approaches based on fuzzy algorithms, artificial neural networks and genetic algorithms [5], [6].

The unclear relations between load and other relevant variables and lack of information have led to the development of these recent methodologies that are much more capable of dealing with the uncertainty [1].

In this paper we develop a method for short-term forecasting based on a Gaussian process (GP) model. This is a probabilistic non-parametric model that provides the prediction and the variance value of the prediction [7]. The prediction value variance that depends on the data density and, noise can be viewed as the prediction level of confidence and is the principal advantage of this method when compared to neural network or fuzzy models. Another advantage of this kind of model is that its structure is determined only by the selection of the covariance function and the regressors, and can be determined even with relatively small training points. A major disadvantage of the Gaussian process model is the computational load associated with the need to invert the covariance matrix, whose dimensions depend on the size of the data set and the number of regressors, at every iteration of the optimization algorithm.

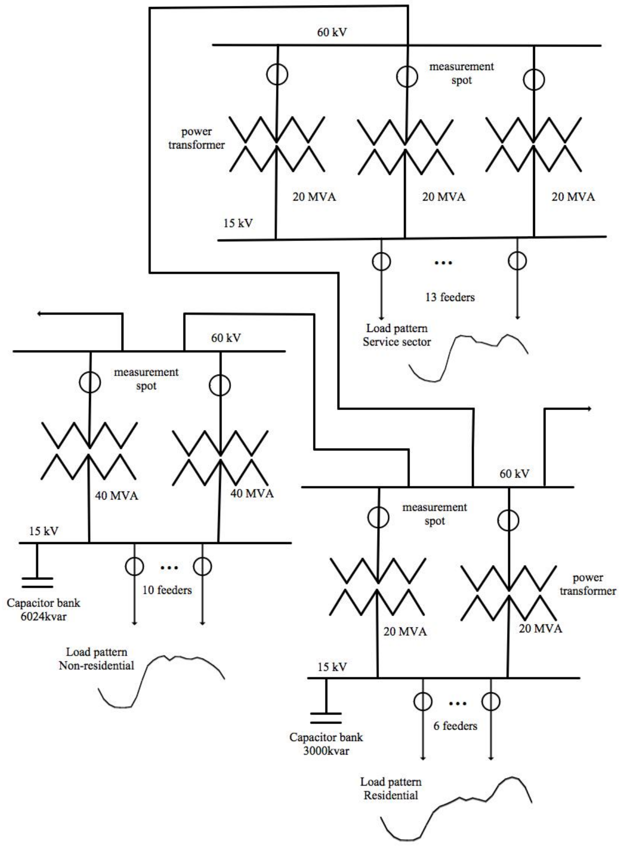

Fig. 1. Simplified on line diagram of the three electrical distribution substations representative of the case study.

This methodology is illustrated by a case study composed of three medium-size supply electrical distribution substations (ES) located in the center of Portugal, near the city of Coimbra. The global scheme is represented in Fig. 1 [8].

The three ES are supplied at 60 kV and the secondary voltage is 15kV, a voltage level that is normal in this type of distribution network. To test the model under different load conditions three different ES were included (Fig.1). These ES have different types of load patterns, namely the residential, non-residential, and a third one representative of the service sector (Fig. 2).

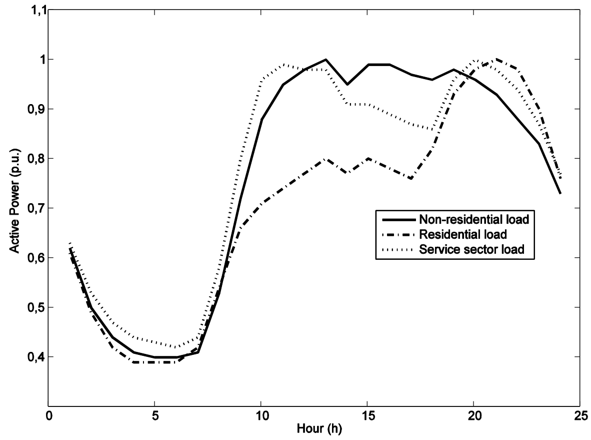

Fig. 2. Different types of load diagrams (p.u.) for the same day (winter).

This last pattern has two typical zones with two peak loads, one in the morning and the other in the evening (Fig.2). This presents an increased difficulty to the forecast model because of the unknown relative position of the two peak loads (morning and evening).

The contributions of the paper consist of a novel methodology for short-term load forecasting, based on Gaussian Processes, and its demonstration in a real-load case study.

The paper is organized in five sections (including this introduction): In section II, the use of Gaussian processes as a method for regression and modeling is briefly presented. In section III, the assembled variables as well as their correlations with the load are analyzed. This section also includes the definition of the structure of the vector of regressors. In section IV, the proposed forecasting method is applied to the electrical load forecast case study. Finally, conclusions are drawn in section V.

II. GAUSSIAN ESTIMATOR PROCESS

A Gaussian process is a collection of random variables, which have a joint multivariate Gaussian distribution. Assuming a relationship of the form *y=f(x)* between an input $\boldsymbol{x} \in \mathbb{R}^D$ and an output $y \in \mathbb{R}$, the output can be viewed as a collection of random variables $y(1), \cdots, y(n) \sim \mathcal{N}(0, \Sigma)$ which have a joint multivariate Gaussian distribution. The covariance matrix Σ can be parameterized and computed by means of a function $\Sigma_{pq} = Cov\big(y(p), y(q)\big) = C\big(\boldsymbol{x}(p), \boldsymbol{x}(q)\big)$ that determines the covariance between output points corresponding to input points $\boldsymbol{x}(p)$ and $\boldsymbol{x}(q)$. The Gaussian process can be fully specified by its mean $\mu(\boldsymbol{x})$ (usually assumed to be zero) and its covariance function $C(\boldsymbol{x}(p),\boldsymbol{x}(q))$. It is remarked that, although not all data can be modeled as a zero-mean process, this assumption is correct if the data is properly scaled and detrended[9].

A common choice of covariance function, when we assume that the process is smooth and stationary (the mean is constant and the covariance function only depends on the distance between the inputs $x(i)$, that has been proven to work well in practice [7] is:

$$C\big(\boldsymbol{x}(p), \boldsymbol{x}(q)\big) = v_1 exp\left[-\frac{1}{2}\sum_{d=1}^{D} \omega_d \big(x^d(p) - x^d(q)\big)\right] + v_0 \delta(p, q) \tag{1}$$

where x^d (p) denotes the d^{th} component of the *D*-dimensional input vector $\boldsymbol{x}(p)$, $\Theta = [v_1, \omega_1, \dots, \omega_D, v_0]^T$ is the vector of hyperparameters and $\delta(p, q)$ is the Kronecker operator defined as:

$$\delta(p, q) = \begin{cases} 1, & p = q \\ 0, & p \neq q \end{cases} \tag{2}$$

The $\omega_1, \dots, \omega_D$ parameters control the scaling of the distances in each input dimension $x^1(i), \dots, x^D(i)$. The parameter ν_1 is the overall scale of correlations and ν_0 expresses the process noise variance. The exponential term suggests that less distant input vectors lead to highly correlated outputs while more distant inputs generate low correlated outputs. The $\omega_1, \dots, \omega_D$ parameters can be used to evaluate the relative importance of the corresponding input components (dimensions), *i.e.*, a high or low ω_i value means that the inputs in dimension *i* contain high or low information, respectively. There are other forms of covariance functions that can be chosen [10], the only restriction is that these covariance functions must generate non-negative definite covariance matrices, for any set of input points.
Assume a statistical model

$$y(k) = f\big(\boldsymbol{x}(k)\big) + \epsilon(k) \tag{3}$$

with an additive uncorrelated Gaussian white noise with variance ν_0, $\epsilon \sim \mathcal{N}(0, v_0)$. Given a set of training data pairs of input data $\boldsymbol{X} = [\boldsymbol{x}(1), \boldsymbol{x}(2), \dots, \boldsymbol{x}(n)]$ and the corresponding vector of output data $\boldsymbol{y} = [y(1), y(2), \dots, y(n)]$ and a Gaussian Process prior on $f(\boldsymbol{x})$, with zero-mean and Gaussian covariance function such as (1) we wish to get the predictive distribution of $y(n+1)$ corresponding to a new given input $\boldsymbol{x}(n+1)$. For the random variables $y(1)$, $y(2)$, , , $y(n)$, $y(n+1)$ we can write:

$$\boldsymbol{y}, y(n+1) \backsim \mathcal{N}(0, \boldsymbol{K}_{n+1}), \tag{4}$$

where $\boldsymbol{K}_{n+1}$ is the covariance matrix made of submatrices as follows:

$$\boldsymbol{K}_{n+1} = \begin{bmatrix} \begin{bmatrix} & \boldsymbol{K} & \end{bmatrix} & \begin{bmatrix} \boldsymbol{k}\big(\boldsymbol{x}(n+1)\big) \end{bmatrix} \\ \begin{bmatrix} \boldsymbol{k}\big(\boldsymbol{x}(n+1)\big)^T \end{bmatrix} & \begin{bmatrix} \mathcal{K}\big(\boldsymbol{x}(n+1)\big) \end{bmatrix} \end{bmatrix} \tag{5}$$

The matrix **K** is the $n \times n$ covariance matrix for the training data such that

$$K = \sum_{pq} = C\left(\boldsymbol{x}(p), \boldsymbol{x}(q)\right) \quad (6)$$

and the vector

$$\boldsymbol{k}\left(\boldsymbol{x}(n+1)\right) = \left[C\left(\boldsymbol{x}(1), \boldsymbol{x}(n+1)\right), \dots, C\left(\boldsymbol{x}(n), \boldsymbol{x}(n+1)\right)\right] \quad (7)$$

is the $n \times 1$ vector of covariances between the training inputs and the new input. The expression

$$\mathcal{K}\left(\boldsymbol{x}(n+1)\right) = C\left(\boldsymbol{x}(n+1), \boldsymbol{x}(n+1)\right) \quad (8)$$

is the autocovariance of the new input.
The conditional distribution of (4) allows to obtain the predictive distribution of $y(n+1)$, which is also Gaussian [11],

$$P\left(y(n+a)/\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{x}(n+1)\right) \sim \mathcal{N}\left(\mu\left(\boldsymbol{x}(n+1)\right), \sigma^2\left(\boldsymbol{x}(n+1)\right)\right) \quad (9)$$

where $\mu\left(\boldsymbol{x}(n+1)\right)$ and $\sigma^2\left(\boldsymbol{x}(n+1)\right)$ are the mean and variance of the Gaussian predictive distribution, and are given by:

$$\mu\left(\boldsymbol{x}(n+1)\right) = \boldsymbol{k}\left(\boldsymbol{x}(n+1)\right)^T \boldsymbol{K}^{-1}\boldsymbol{y} \quad (10)$$

$$\sigma^2\left(\boldsymbol{x}(n+1)\right) = \mathcal{K}\left(\boldsymbol{x}(n+1)\right) - \boldsymbol{k}\left(\boldsymbol{x}(n+1)\right)^T \boldsymbol{K}^{-1}\boldsymbol{k}\left(\boldsymbol{x}(n+1)\right) \quad (11)$$

We may say that, given a new input vector $\boldsymbol{x}(n+1)$, the predicted model output $\hat{y}(n+1)$ is the mean of the Gaussian distribution, *i.e.*, $\hat{y}(n+1) = \mu\left(\boldsymbol{x}(n+1)\right)$ and the uncertainty of this prediction is given by the variance of the Gaussian distribution $\sigma^2\left(\boldsymbol{x}(n+1)\right)$. The predictive mean (10) can be interpreted as a weighted sum of the training outputs $\boldsymbol{y}$, to make a prediction at the test point $\boldsymbol{x}(n+1)$.
To be able to make predictions, based on (10), the vector of hyperparameters Θ has to be provided either as prior knowledge or estimated from the available data. This may be done by maximization of the log-likelihood of the hyperparameters:

$$\begin{aligned} \mathcal{L}(\Theta) &= logP(\boldsymbol{y}/\boldsymbol{X}) \\ &= -\tfrac{1}{2}log(|\boldsymbol{K}|) - \tfrac{1}{2}\boldsymbol{y}^T\boldsymbol{K}^{-1}\boldsymbol{y} - \tfrac{n}{2}\log(2\pi), \end{aligned} \quad (12)$$

Optimization requires the computation of the derivative of $\mathcal{L}(\Theta)$ with respect to each parameter $\omega_1, \dots, \omega_D$, ν_0, ν_1 of the vector of hyperparameters Θ, which is given [12] as

$$\frac{\partial \mathcal{L}(\Theta)}{\partial \Theta_j} = -\frac{1}{2}tr\left[\boldsymbol{K}^{-1}\frac{\partial \boldsymbol{K}}{\partial \Theta_j}\right] + \frac{1}{2}\boldsymbol{y}^T\boldsymbol{K}^{-1}\frac{\partial \boldsymbol{K}}{\partial \Theta_j}\boldsymbol{K}^{-1}\boldsymbol{y}, \quad (13)$$

where *tr* denotes the trace of the matrix.
Gaussian estimator processes can be used for the modeling of dynamic systems of the form (3) if delayed input and output signals are used as regressors [13], [14],

$$\begin{aligned} \boldsymbol{x}(k) = [&y(k-1), y(k-2), \dots, y(k-L), \\ &u(k-1), u(k-2), \dots, u(k-L)] \end{aligned} \quad (14)$$

In such cases an autoregressive model is considered, such that the current output depends on previous outputs *y*, as well as on previous exogenous variables *u*, up to a given lag *L*.
The Gaussian process model not only describes the dynamic characteristics of the system but also provides information about the confidence in the predictions. This advantage can be used to point out predictions of poor quality, indicated by the corresponding variance high values.
For further details, a presentation of Gaussian processes can be found in [15].

III. Data Analysis and pre-processing data

Data analysis was divided in three phases: data preparation, choice of the regressor vector and, finally, an analysis of the week and weekend load patterns.

A.First phase- data preparation

In the first phase the collected data series were pre-processed mainly to fill the gaps caused by failures occurring in the acquisition system (SCADA) operation. For the training, testing and validation of the methodology, over three years of network data were available. Records of this length usually display some occurrences of lost data that have to be reconstructed. The rules for this purpose are: if the lost data period reaches one hour, the average of the previous four 15 minutes values is used; if the lost data period is longer than one hour, the whole day is replaced with the average load profile of two corresponding days from the previous two weeks; if the lost data corresponds to a special day, a custom analysis is performed and the closest similar day or the closest Sunday is used. The hourly time series were constructed based on 15 minutes times series.

Fig. 3 shows the evolution of a typical hourly week load diagram in two different weather situations for the service sector pattern.

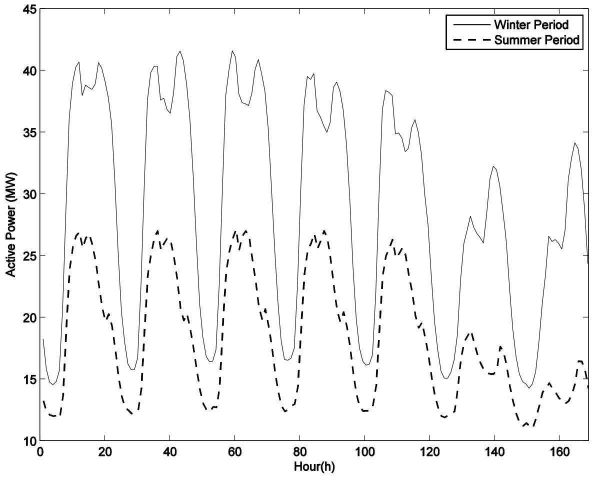

Fig. 3. Typical week load diagram for the same ES in two different climatic situations.

In some regions, under certain climatic conditions, the temperature is strongly correlated to the consumption; particularly in summer when the humidity and temperature are high, or during winter when very low temperatures occur. In this case study, moderate temperature swings are accompanied by moderate humidity conditions [16].

The decrease in daily peak load with the approach of spring and summer reveals a weak penetration of heating ventilation and air-conditioning loads. In the winter, the peak load increases because of the strong presence of heating loads (Fig.3).

In this case study and for the next hour horizon, the correlation between the variable temperature and next hour load is very tenuous. To validate this fact, an exhaustive correlation analysis between hourly load and temperature, with different shifts, was performed [5], [8].

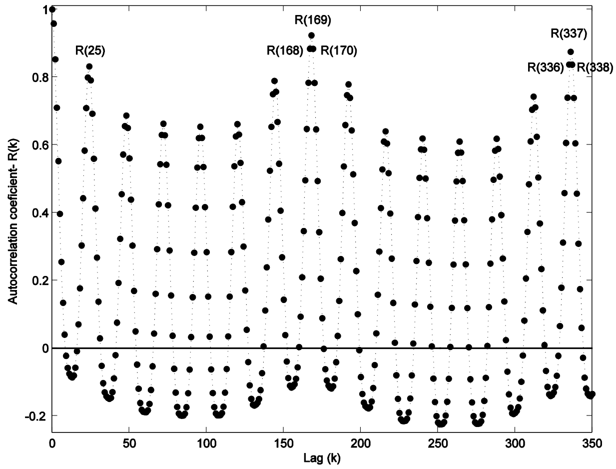

Fig. 4. Autocorrelation coefficients for the active power time-series.

B. Second phase - Choosing the regressor vector

The structure of a GP model is determined by the selection of the covariance function and the regressors, *i.e.*, the components of the input vector $\boldsymbol{x}(t)$. As already mentioned in section II, a common choice for the covariance function is (1). The second part of structure determination comprises the selection of the best subset of regressors in (13) that has the greatest contribution to the output of the non linear model. Incase of (3), this means to select the lags and regressors that should be included in the function *f* argument. The most common method for regressor selection consists of trial and error between several alternatives. The model quality is assessed for each regressor selection until a good result is reached. This method is called *validation based regressor* selection, or *exhaustive search for best regressors* [17], and can be very time-consuming. Other methods can be used, such as the *neighbor methods*, the *estimate and compare methods* or analysis of variance (ANOVA). Surveys regarding these and other approaches for regressor vector selection in non linear models are given in [17] - [19].

A careful analysis was done to reduce the arbitrariness of the regressor selection. The composition of the input variables or vector of regressors was defined through active power autocorrelation coefficients analysis [20]. According to the previous analysis, made in sub section A, load correlations with temperature and relative humidity are weak in the case of the next hour horizon. Therefore, the input vector is only composed of endogenous variables, whose selection is based on the analysis of the autocorrelation coefficients R(k). It may be seen in Fig 4 that the correlations between the most recently available consumption values *y(t)* and those that occurred at past instants y*(t-1) and y(t-2)* are high. The autocorrelation values are also significant for homologous periods, showing relative maxima at two homologous values in the past two adjacent weeks, *R*(*169*) and *R*(*337*), which correspond to the correlation between y(t) and y*(t-168*) and *y(t-336)*, respectively (see Fig. 4).

The autocorrelation evolution allows identifying a certain stationarity of the load time series. One should point out the similarity between the coefficients around the homologous values y*(t-168) and y(t-336)*.

The inclusion of the values *y(t-167)*, *y(t-169)*, *y(t-335)* and *y(t-337)* in the input values provides information regarding the consumption trend in past homologous periods.

The autocorrelation coefficients always decrease as one moves deeper into the past, which can be explained by the seasonal variation of consumption, entailing different load patterns. The final input variables or the vector of regressors for this approach was defined as:

$$\boldsymbol{x}(t) = [y(t-1), y(t-2), y(t-167), y(t-168) \\ y(t-169), y(t-335), y(t-336), y(t-337)] \tag{15}$$

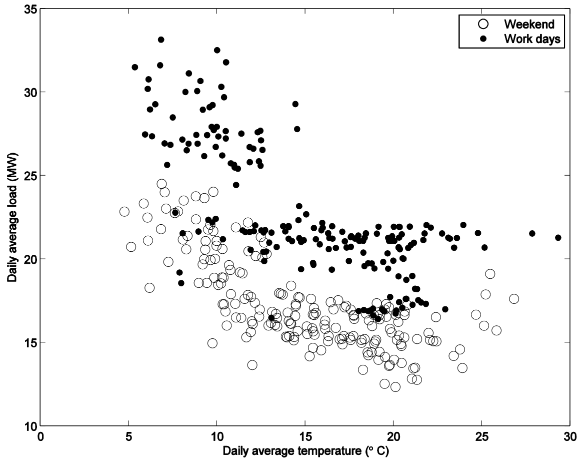

Fig. 5. Average scatter plot between load and temperature (average values).

It is very important to evaluate the relative position of the training data set as regards to the data set used in forecast simulation. If all forecasted values are “contained” in the region defined by the values of the training set, a better performance of the model is to be expected than in situations when the training set of values does not "cover" the values used in the forecast [5].

This lack of “coverage” is natural in this type of signal. Several factors may lead to it, such as management actions like network reconfigurations. This lack of coverage strengthens the need for carrying out careful evaluation of the results and, if such is the case, to develop a new definition of the training set in order to include these effects. This procedure increases the capacity of generalization of the GP model [21].

C. Third phase – Load pattern between weekdays and weekends

The difference between weekdays (see the first 120 hours of Fig. 3) and weekends (see Fig. 3 last 48 hours) is notable in the load patterns. The question arises whether a single model is sufficient to predict the load in these two different situations. In our opinion the answer is affirmative.

If we consider a similar evolution of load in terms of medium average load, and medium day temperature for both weekdays and weekends (Fig. 5), we may conclude that these two different load patterns present the same behavior in terms of average values. During winter the load diagrams present a major peak load caused by the influence of temperature.

The correlation between day load average and day average temperature is expressive (Fig. 5). It is possible to identify two zones in Fig. 5. The first one on the left side of the figure represents a negative correlation coefficient area, and reveals the influence of heating loads. The second one represents an area where the load is independent of temperature. The evolution of load during the weekend and workday is very similar (Fig.5).

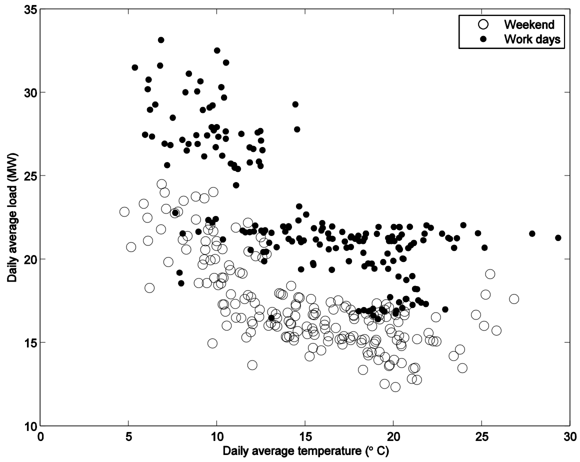

Fig. 5. Average scatter plot between load and temperature (average values).

The input values chosen in sub-section III-B are capable of "explain" the process, for both day types, as confirmed in section IV with the forecast results.

IV. Forecast results - the case study

The GP model was trained, tested and validated for all the ES of the case study, and for extreme climatic situations like the winter and summer periods. The GP model was also tested for all the different load diagrams presented in the case study. The results presented in this sub-section concern the service sector load because the presence of two peak zones, with its inherent uncertainty (Fig.2), brings more complexity to the forecast model.
The GP model was trained with data values obtained from the winter and summer of 2003 and was tested for the summer of 2004 and the winter of 2005. For the training and testing of the GP model the data set was divided into winter and summer periods.
To analyse the quality of the forecast results, the values obtained for some of the most current statistical indicators are analyzed. The most significant indicator, as is generally accepted for comparing different forecast approaches, is the mean absolute percentage error (MAPE) [22], which is defined as follows:

$$MAPE = \sum_{i=1}^{n} \frac{|(y_t - \hat{y}_t)/\hat{y}_t|}{n} \quad (16)$$

where y_t and $\hat{y}_t$ are, respectively, the actual value and the forecast value. Other statistical indicators are also necessary to provide a more comprehensive view of forecast results such as the percentage error (PE) and the mean percentage error (MPE):

$$PE_t = \frac{y_t - \hat{y}_t}{\hat{y}_t} 100 \quad (17)$$

$$MPE = \sum_{t=1}^{n} \frac{PE_t}{n} \quad (18)$$

where n is the size of the sampling in the analysis.
The PE and MPE parameters should not deviate much from zero, as a sign of desirable lack of bias in the forecast series of values.
The forecast results and errors obtained with this approach are in accordance with the expected values obtained with other regression methodologies. In Fig. 6, a typical winter working week is represented. The MAPE value obtained for this period was 1.22%, which is a good result, in line with other approaches [6]. In the distribution sector a MAPE value less than 5% is acceptable. In this case study, and for all the different loadsituations during several periods, the GP presents a significantly lower error level.
In Fig. 7, the forecasts for two consecutive weekends are presented. The MAPE value for these two weekends is 1.26%. The performance of the model is practically the same in these two different situations. The results are in accordance with the analysis made in section III.

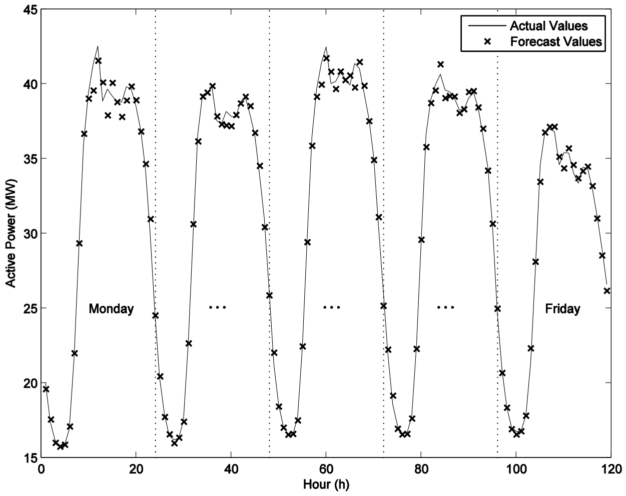

Fig. 6. Working week forecast results in a winter period.

Fig. 8 shows the evolution of the PE for the forecast period represented in Fig. 6, and 7, which includes the weekends.

The time evolution of PE is practically symmetrical around zero with an MPE of 0.05%, which is a sign of desirable lack of bias (Fig. 7).

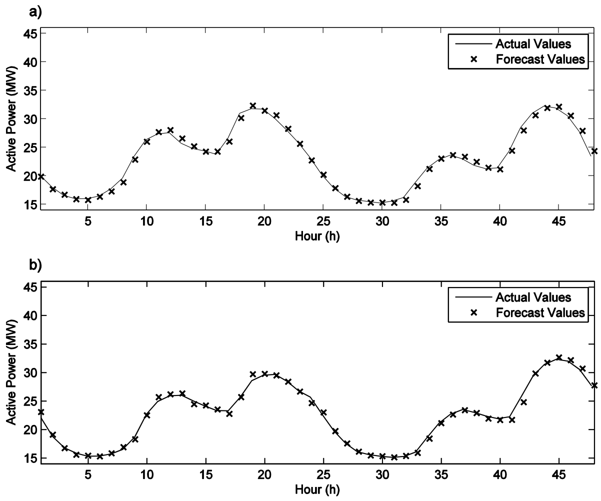

Fig. 7. Forecast results for two consecutive weekends in the winter period: **a)** First weekend, **b)** Second weekend.

For the summer period, the results are better (see Fig. 9). The same load diagram behavior (with a lower average level) allows better MAPE values, 0.92% for working week and 1.33% for weekends, and an MPE of 0.24% for the whole week. The same model is used for both climatic situations.

Another important statistical indicator is the confidence level of the forecast values. Fig. 10 depicts, for the winter period, the actual active power values and the 95% confidence region of the forecasts (shaded area). We can see that most of the actual values, nearly 98% (209 in 216), lie within the confidence region of the forecasts.

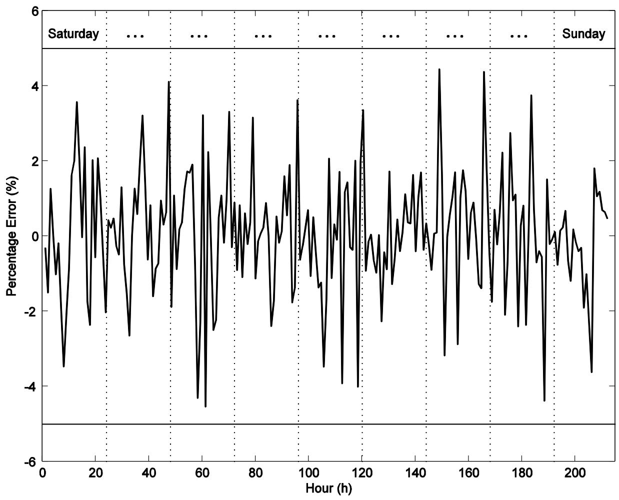

Fig. 8. Percentage error PE in winter period.

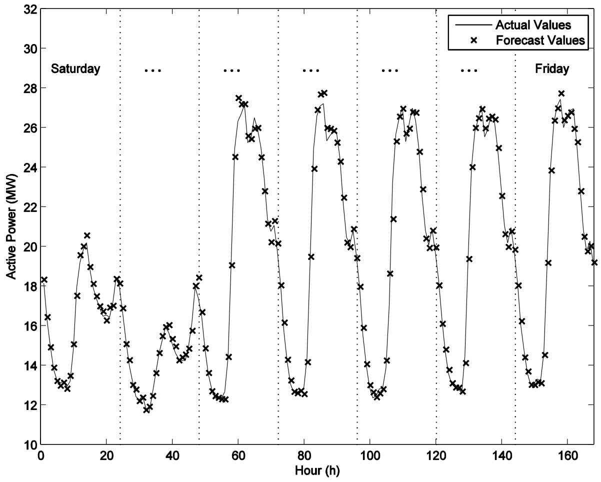

Fig. 9. Forecast results in the summer period.

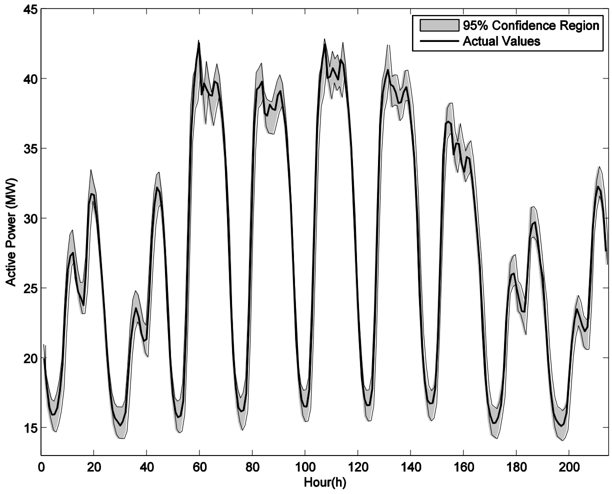

Fig. 10. The 95% confidence interval for the forecast values in the winter period.

V. CONCLUSIONS

The application of short-term methodologies in the distribution sector, particularly in a de-regulated environment, has a significant role in distribution companies. Actions like network management, load dispatch and network reconfiguration, under quality of service constraints, require reliable short-term load forecasts.

This paper presents a Gaussian Process modeling approach as a new STLF methodology applied to the distribution sector. This approach, besides imparting forecasts, also provides information about the confidence in the forecasts. In the modeling procedure, the vector of regressors has to be chosen. Based on correlation coefficient analysis, a careful selection of the regressors was made. These include a short stream of contiguous values prior to the forecast to be made, and past values collected at homologous periods of preceding weeks representing the trend of consumption. The GP model was trained and tested with three years of real-life data values. We have shown that the GP approach can achieve good results. Values of MAPE are consistent and always less than 1.5% for different climatic situations.

Besides the service sector, the GP methodology was also applied to other load consumer patterns such as the residential and the non-residential. For these three types of load the GP model has achieved a good level of accuracy.

The GP approach establishes an interesting approach in the short-term discourse, when applied to the distribution sector, with very satisfactory results.

ACKNOWLEDGMENT

The authors gratefully acknowledge the contributions of EDP Distribuição (Portugal) - Direcção do

Centro for all the provided data.

REFERENCES

[1] G. Gross, F.D. Galiana, "Short term load forecasting," *IEEE Proceedings*, vol. 75, nº 12, Dec. 1987; 1558:1573.

[2] A. Phatchakorn, S. Tomonobu, T. Hirofumi, and Y. Atsushi. "A Hybrid ARIMA and Neural Network model for short term Price Forecasting in Deregulated Market ". *IEEE Trans. on Power Systems*, to be published.

[3] M. Caciotta, R. Lamedica, V. Cencelli, A. Prudenzi, and M. Sforna, "Application of artificial neural networks to historical data analysis for short-term electric load forecasting," *European Trans. On Electrical Power-ETEP*, vol. 7, pp. 49-56, 1997.

[4] C.S. Chen, Y.M. Tzeng, and J.C. Hwang, "The application of artificial neural networks to substation load forecasting," *Electric Power Systems Research- EPSR*, vol. 38, pp. 153-160, Aug.1996.

[5] P.J. Santos, A.G. Martins and A.J.Pires , " Designing the input vector to ANN-based models for short-term load forecast in electricity distribution systems," *International Journal of Electrical Power & Energy Systems*, vol. 29, pp 338-347, May 2007.

[6] H.S. Hippert, C.E. Pedreira, and R.C. Souza, "Neural networks for short-term load forecasting: A review and evaluation," *IEEE Trans. on Power Systems*, vol. 16, pp. 44-55, Feb. 2001.

[7] C. K. I. Williams and C. E. Rasmussen, "Gaussian Processes for regression," *Neural Information Processing Systems*, M. E. Hasselmo Touretzky, M. C. Mozer, Eds, vol 8, pp 514-520. MIT Press, Cambridge, Ma, 1996.

[8] P.J. Santos, A.G. Martins an A.J.Pires, "Next hour load forecast in medium voltage electricity distribution," *International Journal of Energy Sector Management*, vol.2, nº 3, pp 439-448, 2008.

[9] M. N. Gibbs, "Bayesian Gaussian processes for regression and classification," Ph. D. Thesis. Cambridge, UK: Cambridge University,1997.

[10] P. Abrahamsen, "A review of Gaussian random fields and correlation functions," Technical report 917 (2nd ed.). Oslo, Norway: Norwegian Computing Center, 1997.

[11] C. K. I. Williams, "Prediction with Gaussian processes: From linear regression to linear prediction and beyond," Learning in Graphical Models, M. I. Jordan, Edt, pp 599-621. Kluwer Academic, Dordrecht, 1998.

[12] D. j. C. Mackay. (1997). *Gaussian Processes – A Replacement for Supervised Neural Networks* . Lectures notes for a tutorial at NIPS'97 UK.Available: http//www.inference.phy.cam.ac.uk/mackay/BayesGP.html

[13] J. Kocijan, A. Girard, B. Banco, R. Murray-Smith. "Dynamic Systems Identification with Gaussian Processes," *Mathematical and Computer Modeling of dynamical Systems*, vol 11- nº 4, pp 411-424, December 2005.

[14] A. Girard. "Approximate methods for propagation of uncertainty with Gaussian process models," Ph. D. dissertation, University of Glasgow, Glasgow, U.K., 2004.

[15] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, Ma: MIT Press, 2006.

[16] Radwan E, Abdel-Aal. "Short-term hourly load forecasting using abductive networks," *IEEE Trans. on Power Systems*, vol. 19, pp. 164-173, Fev. 2004.

[17] I. Lind and L. Ljung. "Regressor Selection with the Analysis of Variance Method," Automatica, vol 41,- nº 4, pp 693-700, 2005

[18] R. Haber and H. Unbehauen. "Structure identification of nonlinear dynamic systems—a survey on input/output approaches," *Automatica*, vol 26, pp 651–677, 1990.

[19] I. Lind. "Regressor and structure selection: Uses of ANOVA in system identification," Ph.D. Thesis. Linköping, Sweden, Linköpings universitet, 2006.

[20] J.F. Martins, P.J. Santos, A.J. Pires, L.E.B. da Silva, and R.V. Mendes , "Entropy-Based Choise of a Neural Network Drive Model," *IEEE Transactions on Industrial Electronics*, vol. 54, pp 110-116, Feb. 2007.

[21] K. Azman, J. Kocijan, "Application of Gaussian processes for black-box modelling of biosystems," *ISA transactions*, vol 46, pp 443-457, Oct. 2007.

[22] S. De Lurgio, *Forecasting principles and applications.*. Statistics & probability series Singapore: McGraw-Hill International Editions, 1998, pp. 36–58.