

Instituto de Engenharia de Sistemas e Computadores de Coimbra
Institute of Systems Engineering and Computers
INESC - Coimbra

Marisa Resende and Paulo J. Santos

**Short – Term Load Forecasting Using a Gaussian Process Model –
Optimal Endogenous Regressor**

No.10

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 a Gaussian Process Model- Optimal Endogenous Regressor

Marisa Resende⁽¹⁾ and Paulo J. Santos⁽²⁾

August 3, 2010

⁽¹⁾ Institute of Systems Engineering and Computers of Coimbra. Rua Antero de Quental, N°199, 3000-033 Coimbra, Portugal.

E-mail: *marisa_resende@hotmail.com*

⁽²⁾ Department of Electrical Engineering, Polytechnic Institute of Setúbal, IPS.

Institute of Systems Engineering and Computers of Coimbra. Rua Antero de Quental, N°199, 3000-033 Coimbra, Portugal.

Email: *paulo.santos@estsetubal.ips.pt*

Abstract:

The presence of an electrical de-regulated market reinforces the need of forecast. Actions like network management, load dispatch and network reconfiguration under quality of service constraints, require reliable load forecasts.

This paper establishes a methodological approach based on a Gaussian Process Model in order to choose an efficient endogenous information to be included in the input vector of the forecast. The proposed approach was tested on real-load from three medium-sized supply electrical distribution substations located in the center of Portugal. Considering two types of updating (hourly and daily) for the used load information, the aim will be focused on establishing the optimal regressor vector that can produce the best support load diagrams of prediction. The obtained final results are in accordance with the best values of expected errors for these types of methodologies.

The last procedures reveal extremely importance and utility in activities such as planning, operating and controlling electric energy systems.

Keywords: Gaussian Process Model, Electrical Distribution Substations, Load Forecasting, Covariance Function, Regressor Vector.

1 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 de-regulated environment, commercial transactions take place with a reasonable independence of the technical issues of network management [1],[5]. The system operators (SO) have to rely as much as possible on reliable data [6].

In this paper a method for short-term load forecasting based on a Gaussian Process Model (GPM) is developed. This is a probabilistic non-parametric model that provides the prediction and the variance value of the prediction. One of the great advantages of this kind of model is the fact that its structure can be determined only by the selection of the covariance function and the regressors.

The implicit methodology is illustrated by a case study composed of three medium-size supply Electrical Distribution Substations (EDS) located in the center of Portugal, near the city of Coimbra. The three EDS are supplied at 60 kV and the secondary voltage is 15 kV, a voltage level that is normal in this type of distribution network. To test the model under different load conditions, three different EDS were included. These EDS have different types of load patterns, namely the residential, non-residential, and a third one representative of the service sector [10].

The paper is organized in five sections. In the second section, the use of GPM methodology for regression and modeling is presented. In the third section, the structure of the input regressor vector starts to be defined. Furthermore, the assembled variables as well as their correlations with the load are analyzed in order to select the strictly necessary instances endogenous variables. Since they are going to be considered two types of updating (hourly and daily) of the load values, this section is posteriorly divided in two subsections, the first one for the hourly treatment and the second one for the daily. Each subsection is organized in order to establish the optimal components of the regressor vector, picked from the selected ones, using two kinds of methods, the first one using them chronological contiguous ordered and the second one opting by a decreasing correlational contiguous ordering. The algorithm methodology associated for both methods is properly explained in the first subsection and posteriorly applied for the two types of updating, always using the same electrical load forecast case study. A global analysis of the forecast values produced by each method, for each considered updating, is presented in section four. Finally, conclusions are drawn in section five.

2 Methodology

A GPM methodology is used for building the model associated to the available data [4],[7].

It is necessary to define a $n \times d$ training input matrix X , given by $X = [x(1), \dots, x(n)]^T$, where, $x(i) = [x_1(i), \dots, x_d(i)] \in \mathbf{R}^d$, $i = 1, \dots, n$, and the correspondent training output vector Y , $Y = [y_1, \dots, y_n]^T \in \mathbf{R}^n$.

It is adopted a statistical model $Y = f(X) + \varepsilon$, with $\varepsilon \sim N(0, \sigma_0^2 I_n)$, where I_n is the $n \times n$ identity matrix.

Moreover, the assumed GPM prior on $f(X)$ is a smooth and stationary process with zero-mean and covariance matrix $K(X, X)$. This means that the output y_1, \dots, y_n can be viewed as a sequence of random variables having a joint multivariate Gaussian distribution $N(0, K(X, X))$. In this way, $K(X, X)$ is a $n \times n$ matrix, whose entries K_{ij} , $i, j = 1, \dots, n$, are only dependent from the input vectors $x(i)$ and $x(j)$, which turns possible to write:

$$K_{ij} := Cov(y_i, y_j) = C(x(i), x(j)). \quad (1)$$

The only considered choice for the function C is found at [8], with the defined associated expression:

$$C(x(i), x(j)) = \sigma_1^2 \exp\left(-\frac{1}{2}(x(i) - x(j))^T M (x(i) - x(j))\right) + \sigma_0^2 \delta_{ij}, \quad (2)$$

where σ_0^2 is the process noise variance, σ_1^2 is the signal variance, M is a $d \times d$ diagonal matrix, of the form $M = diag(w_1^2, \dots, w_d^2)$, and δ_{ij} is the Kronecker operator defined as:

$$\delta_{ij} = \begin{cases} 1 & , \quad i = j \\ 0 & , \quad i \neq j. \end{cases} \quad (3)$$

The parameters w_1, \dots, w_d are positive and they can determine the two levels of information (high and low) given, respectively, by the input dimensions $x_1(i), \dots, x_d(i)$, $i = 1, \dots, n$.

In particular, $w_1^{-1}, \dots, w_d^{-1}$ are called the characteristic length-scales.

$\Theta = [w_1, \dots, w_d, \sigma_1, \sigma_0]^T$ is called the vector of hyperparameters.

One can notice that if the output vector Y has non-zero mean, it shall be centered around zero, since it is considered a zero-mean gaussian process.

It won't be necessary to proceed for any other re-scaling since the used training data does not present very different standard deviations for the different input dimensions $x_1(i), \dots, x_d(i)$, $i = 1, \dots, n$.

For making predictions, it is used a $m \times d$ test input matrix X^* , given by $X^* = [x^*(1), \dots, x^*(m)]^T$, where $x^*(i) = [x_1^*(i), \dots, x_d^*(i)] \in \mathbf{R}^d$, $i = 1, \dots, m$.

Before predicting the correspondent output vector Y^* , $Y^* = [y_1^*, \dots, y_m^*]^T \in \mathbb{R}^m$, it is necessary to estimate the vector of hyperparameters, Θ , from the available training input data (Y, X) . For that purpose, the values which maximize the log-likelihood function of the hyperparameters have to be found.

The log-likelihood function is given by the follow equation [8]:

$$L(\Theta) := \log P(Y|X) = -\frac{1}{2} \log(|K(X, X)|) - \frac{1}{2} Y^T (K(X, X))^{-1} Y - \frac{n}{2} \log(2\pi), \quad (4)$$

where $|K(X, X)|$ is the determinant of the $n \times n$ covariance matrix, defined in (1) and (2), for the training input data.

Denoting the hyperparameters $w_1, \dots, w_d, \sigma_1, \sigma_0$, respectively, by the parameters $\theta_1, \theta_2, \theta_3, \dots, \theta_{d+2}$, it is also derived, for each $j = 1, \dots, d + 2$ and according to [8], that

$$\frac{\partial L(\Theta)}{\partial \theta_j} = -\frac{1}{2} \text{tr} \left[(K(X, X))^{-1} \frac{\partial K(X, X)}{\partial \theta_j} \right] + \frac{1}{2} Y^T (K(X, X))^{-1} \frac{\partial K(X, X)}{\partial \theta_j} (K(X, X))^{-1} Y, \quad (5)$$

where tr denotes the trace of the matrix in cause.

In order to find the maximum of $L(\Theta)$, it is necessary to calculate the roots of the last expressions, which implies the solving of a non-linear system of $d + 2$ equations in order to the $d + 2$ variables $\theta_1, \dots, \theta_{d+2}$. This operation involves a certain complexity in terms of computing.

To overwhelm this problem, an optimization routine based on [9] was implemented. Starting with an appropriate point, the implicit algorithm develops line searches that can converge to a local minimum of a differentiable multivariate function.

The negative log-likelihood $(-L(\Theta))$ is an example of a function in those last conditions. Consequently, starting with a particular $\Theta^0 = [\theta_1^0, \dots, \theta_d^0, \theta_{d+1}^0, \theta_{d+2}^0]^T$, the mentioned code allows to obtain the estimated hyperparameters $\hat{w}_1, \dots, \hat{w}_d, \hat{\sigma}_1, \hat{\sigma}_0$, which maximize (4).

The attribution at (2) of the last values results on the estimated covariance matrix $\hat{K}(X, X)$, which gives ability to forecast from the given test input points at X^* .

Given X, Y and X^* , the conditional distribution of the test output, Y^* , is Gaussian [8]. Thus, it will be considered its mean, \hat{Y}^* , as the predicted model output (whose components are denominated forecast values) and its variance, $\hat{\sigma}_{Y^*}^2$, as the uncertainty of the prediction.

In accordance with (1) and (2), one can consider the analogous definitions of the $m \times n$, $m \times m$ and $n \times m$, respectively, covariance matrices $\hat{K}(X^*, X)$, $\hat{K}(X^*, X^*)$ and $\hat{K}(X, X^*)$.

Then it is possible to write, by [8], that:

$$\hat{Y}^* = \hat{K}(X^*, X) (\hat{K}(X, X))^{-1} Y \quad (6)$$

$$\hat{\sigma}_{Y^*}^2 = \hat{K}(X^*, X^*) - \hat{K}(X^*, X) (\hat{K}(X, X))^{-1} \hat{K}(X, X^*). \quad (7)$$

Instead of directly inverting the matrix $\hat{K}(X, X)$ at (4), (5), (6) and (7), it will be applied Cholesky factorization for that purpose, since it is faster and numerically more stable.

Knowing the actual values in Y^* , it is possible to assess the accuracy of the obtained forecast values at \hat{Y}^* , by analyzing the vector of the residuals, $\hat{Y}^* - Y^*$, and computing the associate indicators, Mean Squared Error (**MSE**), Mean Percentage Error (**MPE**) and Mean Absolute Percentage Error (**MAPE**), defined, respectively, as:

$$MSE = \frac{1}{m} \sum_{i=1}^m (\hat{Y}_i^* - Y_i^*)^2 \quad (8)$$

$$MPE = \frac{1}{m} \sum_{i=1}^m \frac{\hat{Y}_i^* - Y_i^*}{\hat{Y}_i^*} \times 100\% \quad (9)$$

$$MAPE = \frac{1}{m} \sum_{i=1}^m \left| \frac{\hat{Y}_i^* - Y_i^*}{\hat{Y}_i^*} \right| \times 100\% \quad (10)$$

Each one of the previous parameters can evaluate the deviation of the residuals around zero and so the fitness of the predicted model to the observed data.

One can see that MPE and MAPE have a similar structure, but, comparatively, since MPE does not consider the absolute relative errors, MPE is smaller, providing a better notion about the signal variation of the residuals. Consequently, it is desirable that MPE value does not deviate too much from zero.

In terms of comparing the performance of different forecast approaches, MAPE indicator will be the most relevant for choosing the best model. However, the other two parameters are also crucial to complement the analysis, especially when the associated MAPE values are analogous.

3 Choosing the input regressor vector

In this section it will be considered an autoregressive model, which means that the current output, Y , depends on previous outputs up to a given lag L .

Without loss of generality, the regressor vector, v_k , $k > L$, will have the form,

$$v_k = [y_{k-1}, \dots, y_{k-L}]. \quad (11)$$

Naturally, the length of the regressor vector determines the dimension of the training input vectors of the matrix X .

The aim of this work is to find the optimal regressor vector for an autoregressive model based on the time series of the active powers collected from the service center. One should point out that they are going to be used the available records reported in the year of 2003.

In figure 1 is presented a characteristic load time series evolution for a specific time of the winter period, enclosing the last four weeks of January and the first day of February.

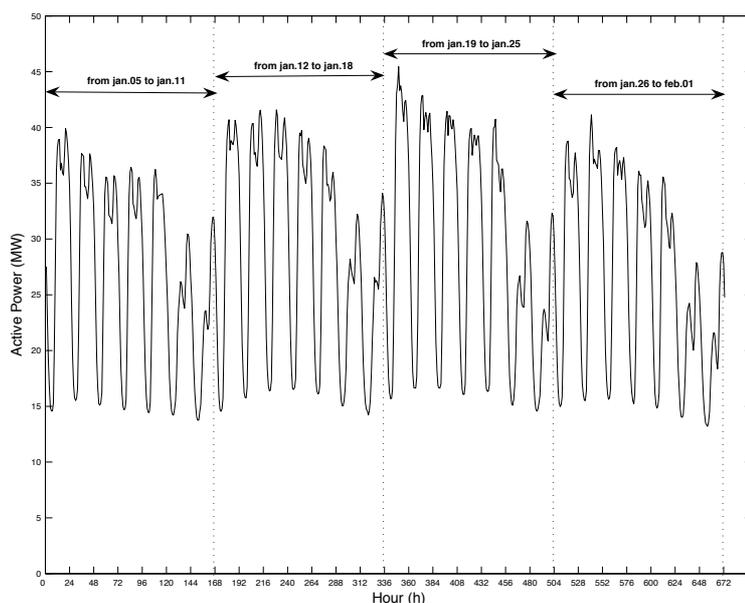


Figure 1: time series evolution of the active powers from 0 : 00 of January, 5, to 0:00 of February, 2.

As one can see, the designated arrows show homologous behaviors in each week, which translates a typical necessary stationarity inherent to a Gaussian Process adapted to the observations.

In order to predict the load in a similar climatic situation, the chosen object of the forecast contemplates the 168 hourly values of the first week of February (from 1 : 00 of Monday, 2, to 0 : 00 of Monday, 9). The establishment of the optimal components for the implied regressor vector can

be improved by the known past information expressed in figure 1. Specifically, for these data, it will be studied the dependence between the registered values of the last week and the observed ones of the previous three weeks, by computing the correspondent autocorrelation coefficients [3].

Denoting by $[y(t)]$ the active power vector, Y , composed by the 168 observed hourly values from 1 : 00 of January, 26, to 0 : 00 of February, 2, it is constructed, for each lag k , $k = 1, \dots, 505$, the correspondent active power vector, $[y(t - k)]$, containing the homologous 168 past hourly values. After that, it is calculated, for each k , the autocorrelation coefficient, $R(k)$, between $y(t)$ and $y(t - k)$. The obtained values are expressed in figure 2, where they had been pointed out all the ones greater than 0.8.

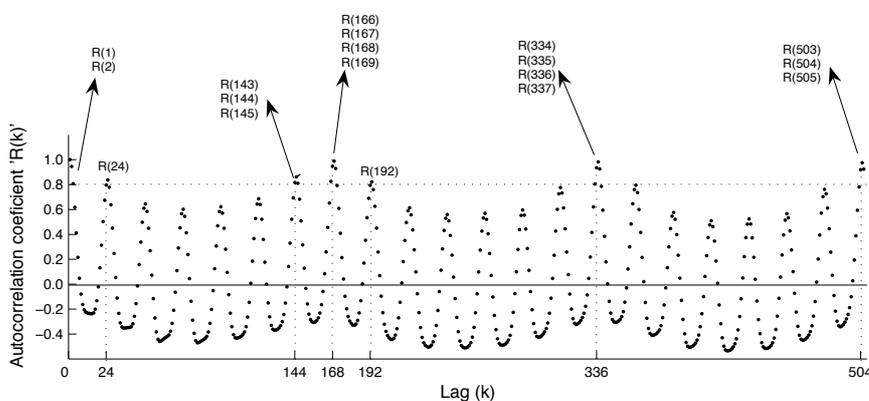


Figure 2: Evolution of the obtained autocorrelation coefficients.

One shall consider that if $R(k)$ is a marked point, then $y(t - k)$ is a possible candidate for a component of the optimal regressor vector. According to figure 2, there are 18 candidates, correspondent to

$$k = 1, 2, 24, 143, 144, 145, 166, 167, 168, 169, 192, 334, 335, 336, 337, 503, 504, 505. \quad (12)$$

The challenge is to find a subgroup that represents the components of the regressor vector which produce the optimal results for the model, that is the lowest MAPE. For this proposal, information about time updating of the observed values is crucial, cause that depends the formulation of the test input.

3.1 Forecasting with hourly updating

This subsection starts with the illustration of a particular application of the GPM, when, for the current output, $Y = [y(t)]$, it is considered the regressor vector $[y(t-1)]$, that constitutes the 168×1 training input vector, X .

In this example, Y has a non-zero mean, 26.3805. As the described methodology is applied for a zero-mean gaussian process, Y has to be centered around zero, which is made by subtracting 26.3805 to each one of its components. Considering Y_c the resultant centered vector, the next step will estimate the vector of the hyperparameters, $\Theta = [w_1, \sigma_1, \sigma_0]^T$, from the training input data (Y_c, X) . To do this, the minimum of the negative log-likelihood has to be found, by applying the appropriate optimization routine mentioned at the last section, starting with $\Theta^0 = [1, 1, \sqrt{0.1}]^T$.

Figure 3 shows the evolution of the 110 developed line searches that assure the convergence to the local minimum, 415.9, of the negative log-likelihood and, simultaneously, the convergence to the values of the hyperparameters, that are going to be used in the new estimated model ($\hat{w}_1 = 55.402$, $\hat{\sigma}_1 = 32.23$, $\hat{\sigma}_0 = 2.729$). These last ones permit compute the new entries of $\hat{K}(X, X)$, attending to (2).

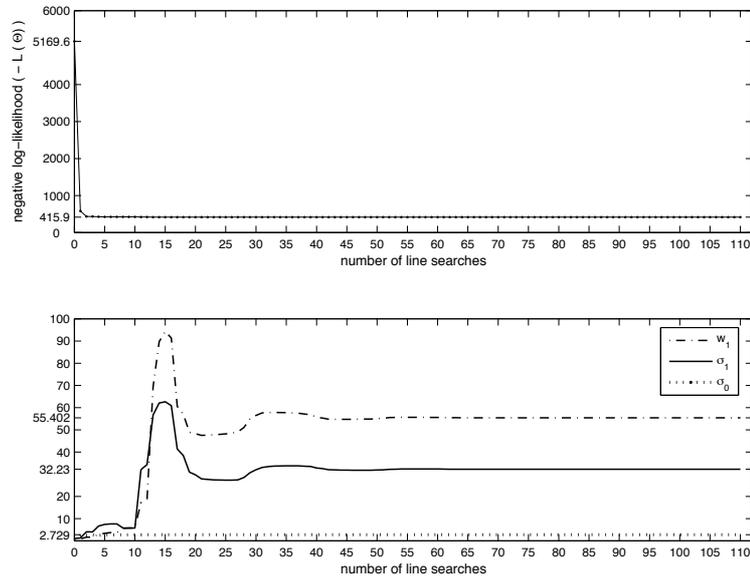


Figure 3: Registered evolution of the produced line-searches for the negative log-likelihood and the hyperparameters.

As it occurs an hourly actualization of the active power values, one can consider the test input column vector, X^* , composed by the 168 hourly values registered from 0 : 00 of Monday, February 2, to 23 : 00 of Sunday, February, 8.

Next, the computing of the new 168×168 covariance matrix $\widehat{K}(X^*, X)$ gives ability to make predictions on the 168×1 output vector, Y_c^* , by applying (6). The forecast values for the model, \widehat{Y}_c^* , are obtained by adding the initial mean value, 26.3805, to each component of the predicted \widehat{Y}_c^* .

The knowledge of the 168×1 actualized column vector, Y^* , and the last obtained predicted column vector, \widehat{Y}_c^* , allows to calculate the associated parameters MAPE, MSE and MPE, whose registered values were, respectively, 8.23%, 6.6188 and 0.17%.

One should note that if for the same output vector, Y , the initial regressor vector had d components, with $d > 1$, the same steps would have to be followed, only using, instead, a $168 \times d$ training input matrix, X , and the initial point $\Theta_0 = [1, \dots, 1, 1, \sqrt{0.1}]^T$, for which the convergence of the method to the minimum of the associated negative log-likelihood would occur.

Posteriorly, with the attainment of the estimated hyperparameters, $\widehat{w}_1, \dots, \widehat{w}_d, \widehat{\sigma}_1, \widehat{\sigma}_0$, and the formulation of the $168 \times d$ test input matrix, X^* , the analogous calculations of the implied predictions would be performed.

Now, the attention is entirely focused on finding the optimal components for the regressor vector. A better performance of the model is expected how more forecast values can be "included" in the region defined by the values of the training input matrix [10].

Next, a possible methodology is presented in order to make the best selection.

One considers the 18 candidates chronological contiguously ordered, from the most recent to the most distant one, according to (12). The strategy consists on starting to take the first one, $y(t-1)$, as a necessary component of the regressor vector, and, step by step, analyzing what next contiguous candidates can be added in order to produce the lowest MAPE.

Specifically, in each step, the contiguous candidate of the last picked one is included in the regressor vector, and, if the resultant MAPE value is smaller than the previous registered one, the added candidate is considered an effective component, otherwise, it is removed.

The last described procedure will be denominated **Chronological Contiguous Choice Method**, or, abbreviating, **CCCM**.

Naturally, besides the MAPE value, the MSE and MPE values are also calculated in each step, in order to get comparisons that can be useful in situations where the produced MAPE values are quite similar.

The obtained results for the considered test input are expressed in figure 4, where, for each step i , $i = 1, \dots, 18$, i^o denotes the considered regressor vector, whose structure progression is specified

above the evolution graphics of MAPE, MSE and MPE.

To simplify notation, the component $y(t - k)$ of each implied regressor vector is simply replaced by k , with k belonging to the set of values discriminate at (12).

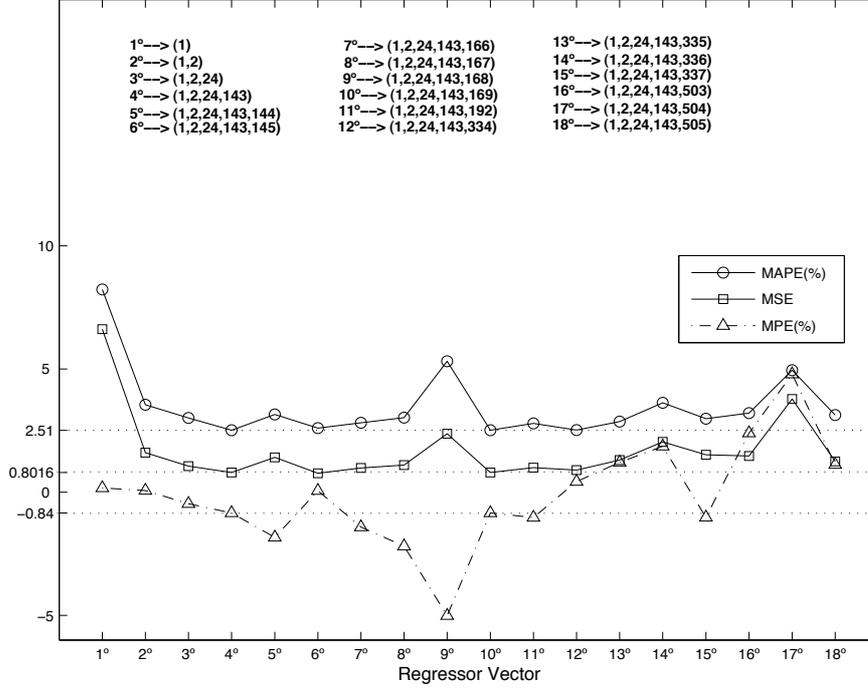


Figure 4: Produced results for the CCCM with hourly updating.

As one can see, till the fourth step a decreasing of MAPE value had took place into 2.51%. In fact, any addition of an available candidate to the regressor vector 4°, never produced MAPE values inferior to the registered 2.51%. It was also indicated 0.8016 and -0.84% , respectively, for the correspondent associated MSE and MPE values.

One should point out that the steps correspondent to 10° and 12° had shown MAPE values almost coincident with 2.51%, but not inferior. On the other hand, for those cases, the associated MSE and absolute MPE values were quite similar to the registered ones with the regressor vector 4°. Thus, $y(t - 169)$ and $y(t - 334)$ were removed, since it is desired an optimal regressor vector with the less possible components. Besides these facts, 6° produced a MAPE value of 2.6%, slightly superior to 2.51%, and, for that reason, did not convince to be a better choice, although its best approximation of the associated MPE value to zero.

In conclusion, using the CCCM,

$$4^\circ \equiv [y(t - 1), y(t - 2), y(t - 24), y(t - 143)] \quad (13)$$

is considered the optimal regressor vector for the model.

Questioning the hypothesis of establishing an optimal regressor vector that can be better than the last one, an analogous method is going to be applied based on the same 18 initial considered candidates, only this time ordered by decreasing of its autocorrelation coefficients, which happens for

$$k = 168, 336, 504, 167, 1, 335, 169, 337, 505, 503, 144, 24, 166, 192, 143, 145, 2, 334. \quad (14)$$

Using exactly the same strategy of the first method, one can found the subgroup of candidates that produces the lowest MAPE.

In this case, the implicit procedure will be denominated **Decreasing Correlational Contiguous Choice Method**, or, abbreviating, **DCCCM**.

The obtained results are the ones expressed in figure 5.

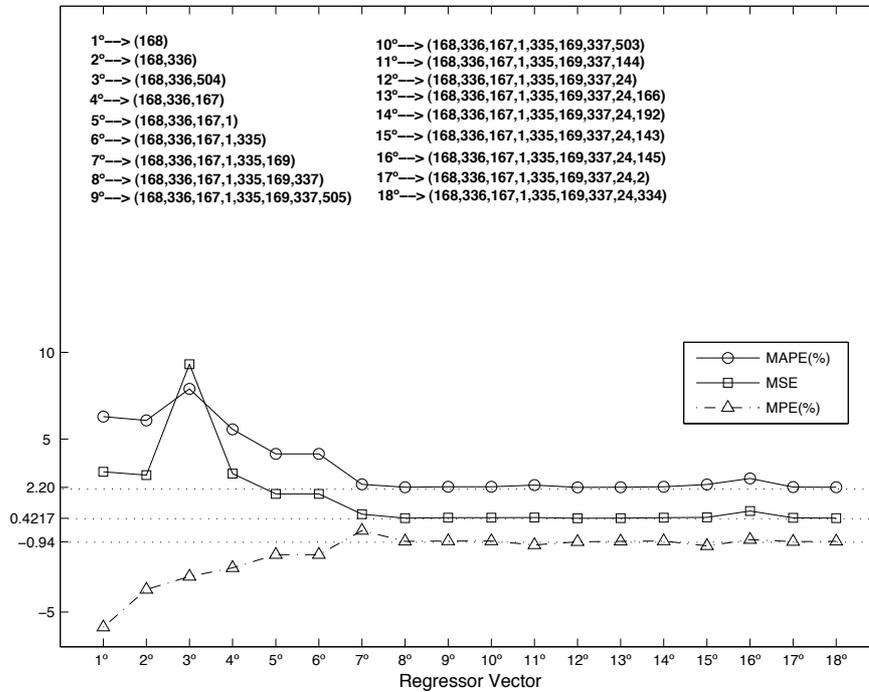


Figure 5: Produced results for the DCCCM with hourly updating.

As the graphic shows, with the exception of the third candidate, $y(t - 504)$, all the other ones, from the first to the eighth, functioned for components of the optimal regressor vector. In spite of the registered similarity of MAPE values for the 5° and 6°, a tiny decreasing effectively took place.

After the eighth regressor vector, the observed similarity for MAPE values was also extended for the associated MSE and MPE values. Consequently, from the eighth step, the final decision only got support by the evaluation of the MAPE progression.

Although it's not very perceptible for the using scale, after starting adding new candidates to the regressor vector 8°, for which the correspondent MAPE value was 2.21%, a slight increasing

of MAPE value had occurred for the next three steps, but a new lowest value, 2.20%, could be obtained with the addition of the tenth second candidate, $y(t - 24)$. After the last inclusion, no more further additions had produced lower values for the MAPE.

In conclusion, using the DCCCM,

$$12^\circ \equiv [y(t - 168), y(t - 336), y(t - 167), y(t - 1), y(t - 335), y(t - 169), y(t - 337), y(t - 24)] \quad (15)$$

is considered the optimal regressor vector for the model.

The comparison of the obtained results by the two described methods shows that, although the DCCCM had presented an inferior MAPE value for the produced optimal regressor vector, the correspondent associated number of components was superior, precisely the double of the registered number for the CCCM.

3.2 Forecasting with daily updating

This subsection will treat the same focused subjects of the last one, but now considering a daily updating of the active power values.

Again, the same 168 hourly values of February, from 1:00 of Monday, 2, to 0:00 of Monday, 9, will constitute the objects of the forecast and it is considered that the load actualization occurs every midnight of each day.

In this type of updating, each hourly load value will be separately predicted, since its forecast can depend from past hourly loads that cannot be already actualized, being used in those cases their estimated value resultant from the performed previous forecasts.

A particular example can illustrate the last description of the method. It will be chosen the same input regressor vector, $[y(t - 1)]$, and output vector, Y , of the last subsection. Consequently, one can obtain the same estimated hyperparameters ($\hat{w}_1 = 55.402, \hat{\sigma}_1 = 32.23, \hat{\sigma}_0 = 2.729$) for the model as well the associated 168×168 covariance matrix $\hat{K}(X, X)$.

The main difference, in relation to the methodology exposed for the hourly updating, stays basically on the formulation of the test input data. Since it is going to be made a prediction every new hour, 168 test input data must be formulated.

For the considered example, each test input, X^* , corresponds to the updated or only estimated previous hourly value for the current time of the forecast. The computation of the associated 1×168 covariance vector, $\hat{K}(X^*, X)$, and the posterior applying of (6) complete the individual prediction. The same process is consecutively repeated till the end.

At 0 : 00 of Monday, February, 9, the updated, Y^* , and predicted, \hat{Y}^* , 168×1 column vectors can be formed and the associated parameters, MAPE, MSE and MPE, calculated. The correspondent obtained values were, respectively, 27.58%, 58.6724 and -1.94% .

Considering the absolute value of MPE, one realizes that the produced results were much higher than the ones obtained for the hourly actualization, and, consequently, the associated model was not so good. This was expected since, between consecutive occurrences of the present updating, the implied algorithm had used twenty three estimated values that could be more or less far way from the reality.

To improve the performance of the forecast, one knows that, for this type of updating, more components must be used in the regressor vector, specially the ones presenting past values with more chances to be actualized. That certainly happens for the ones being not less than twenty four hours behind relatively to each actual time of the prediction.

Consequently, for the daily updating, it is expected that applications of the CCCM and DCCCM can produce a bigger number of components for the associated optimal regressor vectors than the respective ones obtained for the hourly updating.

The next results will show the last suspicion.

To start, it is applied the CCCM. The evolutions of MAPE, MSE and MPE values of the associated algorithm are shown in figure 6.

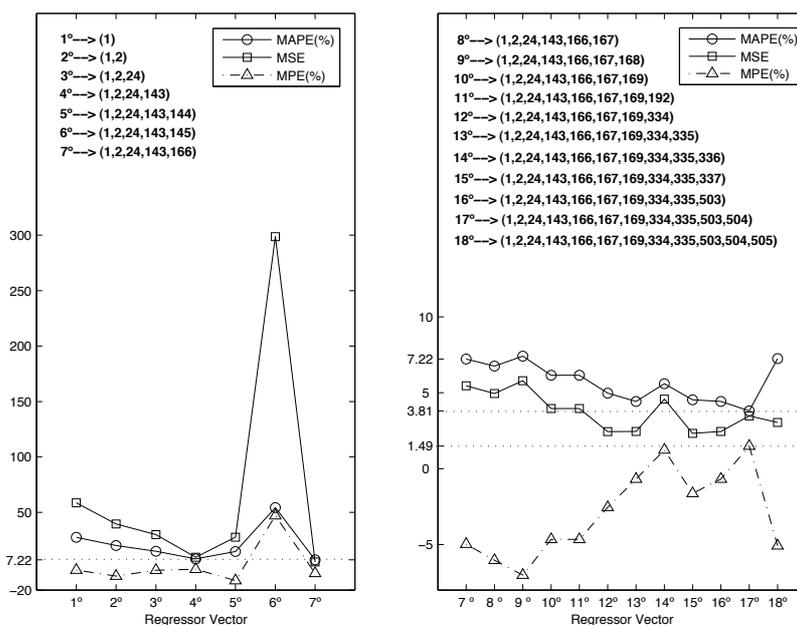


Figure 6: Produced results for the CCCM with daily updating.

Since at the sixth step was registered a surprising but only momentary increasing of MSE value out of the reach of the usual range, the full representation was divided in two others, each one with suitable scales. This way, moving forward from the seventh step, becomes more clearly the particular progress of MAPE and MSE values, due to their variation in a tighter interval.

Clearly, for this method, the optimal regressor vector occurred at the penultimate step, where was registered, for the values of the statistical error indicators, a minimum MAPE (3.81%), a low MSE (3.4739) and a MPE (1.49%) which, in spite of not being too low, it can be included in the registered group of values less distant from zero.

Thus, using the CCCM,

$$17^{\circ} \equiv [y(t-1), y(t-2), y(t-24), y(t-143), y(t-166), y(t-167), y(t-169), \\ y(t-334), y(t-335), y(t-503), y(t-504)] \quad (16)$$

is considered the optimal regressor vector for the model.

Next, it is applied the DCCCM. Figure 7 shows the associated progress of the MAPE, MSE and MPE values.

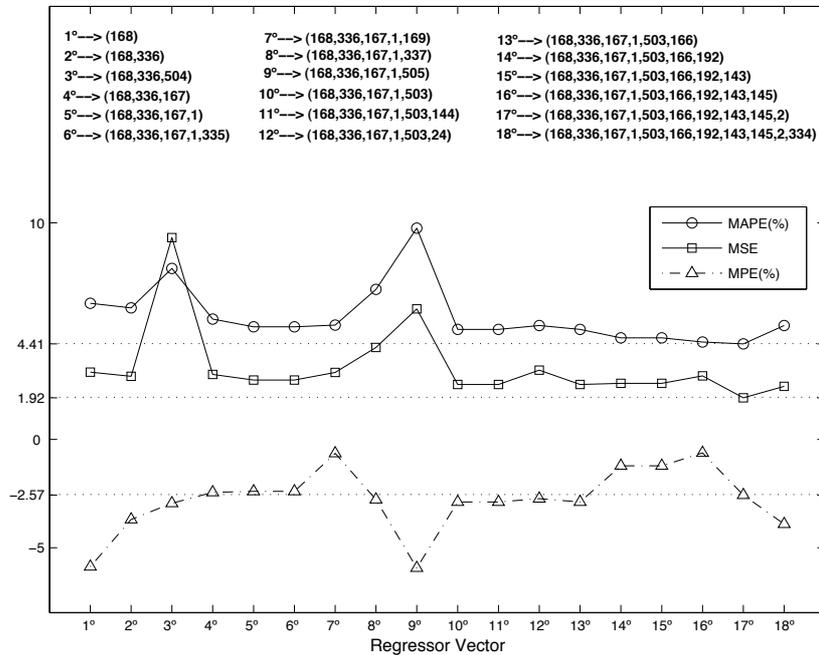


Figure 7: Produced results for the DCCCM with daily updating.

As in the first method, the optimal results were only reached at the penultimate step, only this time the associated regressor vector had registered a less one component.

For that step, MAPE and MSE had presented the lowest values (respectively 4.41% and 1.92), MPE value (-2.57%) was not too good, but still, one of the registered values less moved away from zero. Nevertheless, the produced MAPE was inferior to 5% and the associated MSE low, which had allowed to get a good model anyway.

In conclusion, using the DCCCM,

$$17^\circ \equiv [y(t - 168), y(t - 336), y(t - 167), y(t - 1), y(t - 503), y(t - 166), y(t - 192), \\ y(t - 143), y(t - 145), y(t - 2)] \quad (17)$$

is considered the optimal regressor vector for the model.

4 Overview at the performed forecasts

In order to establish the final considerations towards the obtained results, figure 8 provides a global panorama about the calculated forecast values, for each type of updating (hourly or daily) and used method (CCCM or DCCCM), confronting them with the fixed known actual values. The affected optimal regressor vectors and the MAPE, MSE, MPE values are also discriminated.

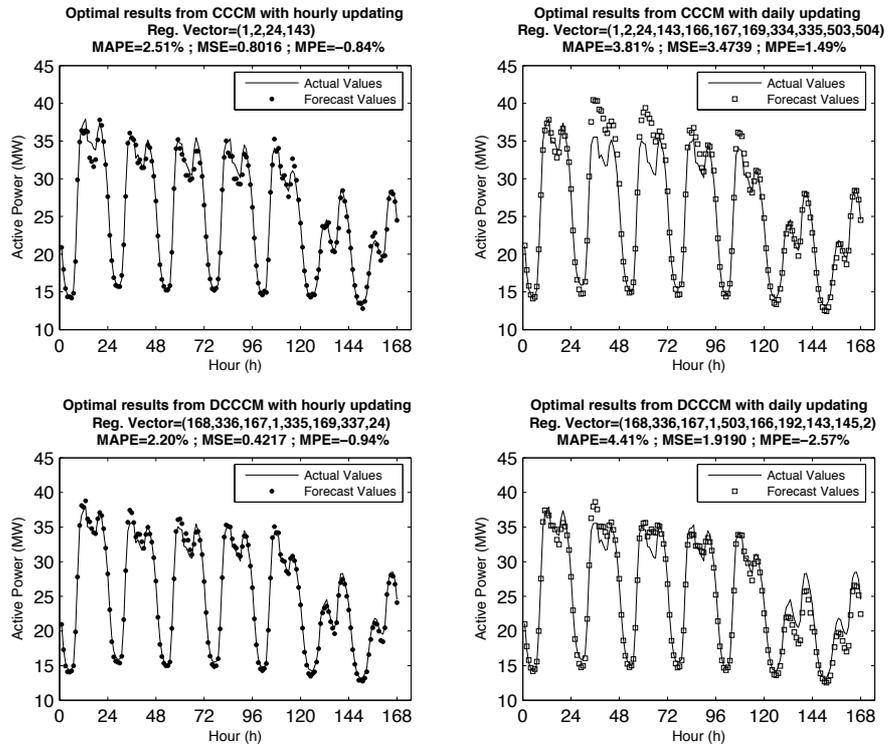


Figure 8: Forecast results for the optimal regressor vectors obtained by the CCCM and DCCCM for the two types of updating.

Naturally, one evidences that, for a fixed type of method (CCCM or DCCCM), the hourly updating could produce better adjusted optimal predictions to the actual values than the daily one. That observation is reflected by the increasing of MAPE, MSE and absolute MPE values, when passing from the hourly to the daily updating.

On the other hand, for a fixed type of updating (hourly or daily), the DCCCM showed, generically, better adjustments of the forecast series of values than the presented ones by the CCCM. This reveals the importance of considering the support of the best autocorrelated candidates in the choice of the best optimal components for the regressor vector.

5 Conclusions

The presented regressive methodology establishes an interesting approach in the short-term load forecasting discourse and evidences a good and consistent error level, for each considered updating. This paper also contributes to reduce a certain arbitrariness in the choice of the best input regressor vector, since its components are picked from a selected group of candidates created from the analysis of the autocorrelation coefficients.

Considering a contiguous chronological or a decreasing correlational ordering of the previous candidates, the respective application of two new explored methods (CCCM and DCCCM) always had produced good results for the case study, since the associated MAPE values always had been lower than the maximum acceptable limit level for the distribution sector (5%).

Naturally, for the hourly updating, lower error levels were obtained. Nevertheless, the satisfactory results achieved for the daily actualization constitute new motivations to future studies based, for example, on the prediction of daily peak load occurrences, with the assumption that data stationarity is a prior condition.

Supported on a fixed small case study, this paper also permit establishing new basis of understanding behavior for the GPM when they are involved training and test load data of higher dimensions.

References

- [1] CACIOTTA, M., LAMEDICA, R., CENCELLI, V., PRUDENZI, A. AND SFORNA, M. (1997), *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.
- [2] DE LURGIO, S. (1998), *Forecasting principles and applications*. Statistics & Probability Series Singapore: Mc-Graw Hill International Editions, pp. 36-58.
- [3] GIBBS, M. N. (1997), *Bayesian Gaussian processes for regression and classification*. Ph. D. Thesis. Cambridge, UK: Cambridge University.
- [4] GIRARD, A. (2004), *Approximate methods for propagation of uncertainty with Gaussian process models*. Ph. D. dissertation, University of Glasgow, Glasgow, U.K.
- [5] GROSS, G. AND GALIANA, F. D. (Dec. 1987), *Short term load forecasting*. IEEE Proceedings, vol. 75, n°12, 1558:1573.
- [6] HIPPERT, H. S., PEDREIRA, C. E. AND SOUZA, R. C. (Feb. 2001), *Neural networks for short-term load forecasting: A review and evaluation*. IEEE Trans. on Power Systems, vol. 16, pp 44-55.
- [7] KOCIJAN, J., GIRARD, A., BANCO, B. AND MURRAY-SMITH, R. (Dec. 2005), *Dynamic Systems Identification with Gaussian Processes*. Mathematical and Computer Modeling of Dynamical Systems, vol 11- n°4, pp. 411-424.
- [8] RASMUSSEN, C. E. AND WILLIAMS, C. K. I. (2006), *Gaussian Processes for Machine Learning*. The MIT Press. www.gaussianprocess.org/gpml.
- [9] RASMUSSEN, C. E. AND WILLIAMS, C. K. I. (2006), *Documentation for GPML Matlab Code*, <http://www.gaussianprocess.org/gpml/code/matlab/doc/>. Updated on June 26th, 2007.
- [10] SANTOS, P. J., MARTINS, A. G. AND PIRES, A. J. (2007), *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.