Proceedings of the American Control Conference
Anchorage, AK May 8-10, 2002

Prediction of Electric Load using Kohonen Maps - Application to the Polish Electricity Consumption

A. Lendasse[1], M. Cottrell[2], V. Wertz[1], M. Verleysen[3]

[1]Université catholique de Louvain, CESAME, 4 av. G. Lemaître,
B-1348 Louvain-la-Neuve, Belgium, {lendasse, wertz}@auto.ucl.ac.be.
[2]Université de Paris I – SAMOS, 90 rue de Tolbiac,
F-75634 Paris Cedex 13, France, Marie.Cottrell@univ-paris1.fr.
[3]Université catholique de Louvain, Electricity Dept., 3 pl. du Levant,
B-1348 Louvain-la-Neuve, Belgium, verleysen@dice.ucl.ac.be.

Abstract. The problem of electrical load forecasting presents some particularities, compared to the generic problem of time-series prediction. One of these particularities is that several values (corresponding to one day of consumption) are usually expected as the result of the prediction. In this paper, we propose an original method dividing the problem into three parts: prediction of the daily mean, of the daily standard deviation and of the normalized daily profile. For the mean and the standard deviation, radial function networks are used as nonlinear approximators. For the normalized profile, a method based on Kohonen maps is proposed. This method is applied to the prediction of the Polish electricity consumption.

1. Introduction

The problem of time series prediction consists in the forecasting of the next value of a series known at consecutive time steps. For example, one could try to predict financial market indexes, the number of spots on the sun surface, the degree of pollution in a city, etc. The prediction of the electric load is specific. For each day, we have 24 values (or more) of the electricity consumption. The problem is the prediction of the next 24 values of the time series i.e. the electric load of the next day. This problem is significantly more difficult; either the prediction is repeated 24 times or the 24 values are predicted simultaneously.

A generic model of time series prediction is presented in Section 2. The use of such a generic method to predict the 24 next values of electric load gives very bad results, because of the different time periods (annual, daily, etc.) encountered in electrical load series. We thus need an original method, more suited to this kind of problem. The solution we propose here is the splitting of the global problem into three simpler ones. The daily profile is characterized by three variables: the mean, the standard deviation and a normalized profile. Each of these three terms will be predicted independently. This general scheme is described in Section 3. The prediction of the mean and the standard deviation is done using radial basis function networks (RBF). These networks are presented in Section 4. To predict the normalized profile, we introduce a particular class of artificial neural networks, the Kohonen maps. These networks are described in Section 5 and their use for the prediction in Section 6. In Section 7, this method is applied to the prediction of the Polish electricity consumption. Although in some cases exogenous data (e.g. the daily temperature) can be used to improve the prediction of the electric load, we do not consider exogenous variable in the sequel.

2. Time Series Prediction

In this section, we will briefly describe a general forecasting method for this method, without exogenous variables [1,2,3,4,5]. We denote the series y_t, with t varying between 1 and n. A general notation for the dynamics of the process is:

$$\hat{y}_{t+1} = f(y_t, y_{t-1}, \dots, y_{t-n}, \theta) , \qquad (1)$$

where θ is the set of parameters that makes it possible for the model f to approximate as well as possible the series. For example, in a Multi-Layer Perceptron (MLP), θ is the set of synaptic weights [6,7]. For RBF, θ will be described in details in Section 4. The vector y_t to y_{t-n} is called the *regressor*. It is obvious that the choice of the regressor and thus of n is capital. If this choice is badly done, the model will be vague or possibly biased. Several methods exist to choose the regressor. For example, one can use the optimal regressor obtained from a linear model. One can also use pruning methods [8], but these usually require extensive computations and are in most cases limited to a particular model. Another technique for the determination of the regressor based on nonlinear projection is presented in [4].

Generally, a model is parameterised by a given number of parameters say M, the size of θ in (1). Usually, when M is small, the model is not complex enough to capture the dynamics of the true system and in case of time series prediction, the prediction will not be accurate. On the contrary, if M is taken too large, the parameters also capture the noise contained in the learning data. This is the overfitting phenomenon. The prediction of the learning data will hence be very accurate (even the noise is "correctly" predicted) but on test data (the generalisation step) the model will be inaccurate. The goal is thus to determine the optimal number M of parameters. With this aim in view, the data y_t will be divided into a learning set LS and a validation set VS. Two different Mean Squared Errors are calculated: the Learning Mean Squared Error (LMSE)

$$LMSE = \frac{\sum_{t \in LS} (\hat{y}_t - y_t)^2}{N_1}, \tag{2}$$

with N_l the dimension of the learning set, and the validation Mean Squared Error (VMSE)

$$VMSE = \frac{\sum_{t \in VS} (\hat{y}_t - y_t)^2}{N_2}, \tag{3}$$

with N_2 the dimension of the validation set. The optimal number of parameters M* that we will choose in our method is the one which gives the minimum VMSE error, while keeping the model parsimonious (with a small number of parameters).

3. General Method

As mentioned in the introduction, the prediction of the next 24 values of the electric load is a difficult problem. A solution to this problem has been developed in [9]. The fundamental steps of our method are based on this solution. The data are denoted $y_{t,i}$ with index t representing the day and index i representing the hour (between 1 and 24). First, we calculate the mean μ_t and the standard deviation S_t for each day of the data:

$$\mu_t = \frac{\sum_{i=1}^{24} y_{t,i}}{24}, \; i = 1,24 \tag{4}$$

$$S_t = \sqrt{\frac{\sum_{i=1}^{24} (y_{t,i} - \mu_t)^2}{24}}. \tag{5}$$

The normalized profile $ny_{t,i}$ can then be obtained for each day using:

$$ny_{t,i} = \frac{y_{t,i} - \mu_t}{S_t}. \tag{6}$$

If predictions of μ_t, S_t and $ny_{t,i}$ are reliable, the prediction of the next 24 hours becomes:

$$\hat{y}_{t,i} = n\hat{y}_{t,i} \cdot \hat{S}_t + \hat{\mu}_t, \; i = 1,24. \tag{7}$$

The prediction of μ_t or S_t is a classical problem of univariate time series prediction. The nonlinear model used for this purpose is the RBF network presented in Section 4. The prediction of $ny_{t,i}$ will be detailed in Section 6.

4. Radial Basis Functions Networks

Given a set of inputs x_i and a set of outputs y_i, approximated by $\hat{y}_i$. The approximation $\hat{y}_i$.of y_i with a RBF model [10,11] is a weighted sum of m Gaussian functions Φ:

$$\hat{y}_i = \sum_{j=1}^{m} \lambda_j \Phi(x_i, C_j, \sigma_j), \tag{8}$$

$$\text{with } \Phi(x_i, C_j, \sigma_j) = e^{-\left(\frac{\|x_i - C_j\|}{\sqrt{2}\sigma_j}\right)^2}. \tag{9}$$

The complexity of the RBF is related to the number of Gaussian kernels. The parameters that need to be determined are the position of the Gaussian kernels (C_j), their widths (σ_j) and the weight factors (λ_j). The technique used to determine these parameters is developed in details in [12]. The following paragraphs summarize this technique.

The positions of the Gaussian kernels are chosen as a function of the distribution of x_i in the input space. Where the density of inputs x_i is low, is placed a small number of kernels; conversely, where the density is high, is used a larger number of kernels. The technique performing this operation is named vector quantization; the points that summarize the initial data (i.e. the centers of the kernels) are named centroids. Vector quantization consists in two steps. First, centroids are initialized at random in the input space. Then, for each input x_i, the closest centroid is moved in the direction of x_i using the following formula:

$$C_j := C_j + \alpha(x_i - C_j), \tag{10}$$

with x_j the considered point, C_j the centroid closest to x_i and α a parameter that decreases with time. The operation is repeated for all data x_i. Further details on these quantization methods can be found in [13].

The second set of parameters that need to be determined is the widths σ_j of the Gaussian kernels. We chose to work with a different width for each kernel. For this purpose, we define the Voronoi zone of a centroid as the area that is closer to this centroid (center of Gaussian kernel) than to any other one. In each Voronoï zone, the standard deviation of the data is calculated. The width of the corresponding Gaussian kernel is the product of this standard deviation by a factor k. We will choose a value of the factor k that minimizes the VMSE defined in (3). This method for determining σ_j has many advantages; the most important of these advantages is that different Gaussian kernels adequately cover the input space of the RBF.

The last parameters to determine are the weight factors λ_j. As all other parameters are already fixed, they are simply determined by the least squares solution of a system of linear equations (8). The total number of parameters is equal to $m*(n+1)+1$ where n is the dimension of the input space and m the number of Gaussian kernels used in the RBF.

5. Kohonen Maps

Kohonen maps are a particular class of Artificial Neural Networks [14]. A Kohonen map is formed by a group of centroids that are linked together. Usually the map is a two-dimensional rectangle and the centroids are placed on a rectangular lattice. These centroids are numbered with respect to their position in the lattice.

Kohonen maps perform two operations. First, they achieve a vector quantization similar to the one presented in Section 3. In addition, the centroids are positioned in the input space in a way that they form a two-dimensional grid. The centroids are numbered according to their location in this grid. Kohonen maps preserve the topology: after learning, two points in the initial dataset that are close in the input space will be projected either on the same centroid (as in classical VQ), either on different centroids that have neighbouring locations in the grid.

The topology-preservation property if often used for visualization purposes: the two-dimensional grid can be viewed as a two-dimensional space where data are projected by Kohonen's algorithm. In fact the data are not projected on the grid, but rather on the centroids themselves; if a continuous projection is needed, interpolation schemes between centroids may be used.

6. Prediction of Normalized Profiles

The normalized profiles $ny_{t,i}$ are 24-dimensional vectors obtained from (6). A Kohonen map is positioned in this 24-dimensional space. For reasons of simplicity, the Kohonen maps we are going to use are square; they are thus lattices of m by m centroids. To visualize the centroids obtained after learning of the map, each node of the grid (corresponding to a 24-dimensional centroid) is represented as a 24-points curve. As the 24 coordinated of each vector are sequentially organized in time in our application, this kind of representation is a natural way to illustrate a (normalized) daily electrical load.

The square lattice is represented and at each node of the lattice the corresponding 24-dimensional centroid is drawn. (See Figure 10 for an example of this representation).

The centroids that are obtained have two interesting properties. First, they summarize correctly the initial input. Secondly, two centroids that are close in the lattice are also close in the input space; their representations are thus very similar.

For each profile $ny_{t,i}$ of a specific Voronoi zone A, the profile $ny_{t+1,i}$ (i.e. the following normalized profile in the time series) is determined. Next, the Voronoi zone B of this normalized profile $ny_{t+1,i}$ is searched. We will say that a transition exists between these two Voronoi zones. The probability transition between a Voronoi zone A and another zone B is the ratio between the number of transition from A to B, and, the number of data in A. The total probability transitions between the Voronoi zones are calculated using the learning set. The probability transitions between a specific Voronoi zone and the other centroids of a Kohonen map are shown in Figure 11.

Once the Kohonen map is built on the normalized profiles $ny_{t,i}$ and probability transitions for each Voronoi zone calculated, the prediction of a normalized profile $ny_{t+1,i}$ is the result of choosing the most probably transition whose origin is Voronoi zone A including $ny_{t,i}$. The only parameter that has to be determined in this method is the number of centroids (m by m) in the grid. The parameter m that minimizes the VSME is chosen.

7. Polish Electricity Consumption

The series studied in this paper represents the electrical consumption in Poland during 2500 days [9] in the 90s. Unfortunately, the database has been previously scaled so that the units of the electricity consumption are unknown.

The mean μ_t and the standard deviation S_t for each day of the data are computed according to (4) and (5). They are represented in Figures 2 and 3. The data are divided into a learning set (1500 points) and a validation set (1000 points). To determine the size of the regressor, linear models of increasing regressor size are built. According to (3) the VMSE of these models is calculated and the results are shown in Figure 3 for the prediction of μ_t and Figure 4 for the prediction of S_t. In Figure 3, the VMSE obtained with a linear model has a bend for a regressor size equal to 8; this value of the regressor is chosen for the prediction of μ_t.

Similarly, according to Figure 4, a regressor size equal to 7 is chosen for the prediction of S_t.

Thereafter, RBF networks are used to improve the prediction of both μ_t and S_t, while keeping the respective regressor sizes. For the prediction of μ_t, the best RBF has 90 centroids and a factor k equal to 7.5. The values of the VMSE with respect to the number of centroids and to parameter k are shown respectively in Figures 5 and 6. Similarly, for the prediction of S_t, the best RBF has 70 centroids and a factor k equal to 2.25. The values of the VMSE with respect to the number of centroids and to parameter k are shown respectively in Figures 7 and 8. The improvement obtained with nonlinear models gives a decrease of the VMSE of 20% for the prediction of μ_t and 24% for the prediction of S_t.

For the prediction of the profile, the method described in Section 5 is performed with an increasing number of centroids in the Kohonen map. The value of the VMSE with respect to the number centroids is shown in Figure 9 (note that the size used in Figure 9 is the number of centroids in each direction of the map, so that the total number of centroids is the square of this size). The optimal size for the Kohonen map is equal to 8. The 8x8 map obtained after convergence of the Kohonen algorithm on the profiles (6) is represented in Figure 10. In Figure 11, we illustrate on the map, the probability transition between a profile and the next ones.

Finally, we illustrate the daily profiles and their predictions in Figure 12 with a example picked up at random in the validation set. The real value of the normalized consumption is given by the solid line and its prediction corresponds to the dashed line.

This approximation of the profile is very accurate and can be combined to the approximation of μ_t and S_t to obtained the complete prediction using (7). The VMSE obtained by the whole procedure is low in comparison to forecastings obtained on the same data with traditional models.

8. Conclusion

In this paper, we considered the problem of forecasting the electrical load, and more particularly the 24 hourly electrical load values of the next day. The solution we propose consists in splitting the prediction of electric load problem into three simpler ones: univariate non-linear prediction of the daily mean and standard deviation, and multivariate prediction of the normalized daily profile. RBF neural networks are used for the two first predictions, and Kohonen maps for the third one. The use of this method avoids the need of re-using predicted values as inputs for a next forecasting, and therefore improves the accuracy of the prediction when several next values (here 24) are to be predicted.

This paper shows the result obtained on the Polish electricity consumption. Similar results have been obtained on large companies electrical consumption, but cannot be illustrated for confidentiality reasons.

Further work consists in taking into account exogenous variables (like temperatures or the day of the week) in the prediction model of normalized profiles.

Acknowledgement

Michel Verleysen is Research Associate of the Belgian National Fund for Scientific Research (FNRS). Part of this work has been funded by the "Ministère de la Région Wallonne", under the "Programme de Formation et d'Impulsion à la recherche scientifique et technologique". The scientific responsibility rests with its authors.

References

[1] G.E.P. Box, G. Jenkins, Time Series analysis: Forecasting and Control (Cambridge University Press, 1976).

[2] L. Ljung, System Identification - Theory for User (Prentice-Hall, 1987).

[3] A. S. Weigend, N. A. Gershenfeld, Times Series Prediction: Forecasting the future and Understanding the Past (Addison-Wesley Publishing Company, 1994).

[4] A. Lendasse, E. de Bodt, V. Wertz and M. Verleysen, Nonlinear financial time series forecasting – Application to the Bel20 stock market index, European Journal of Economic and Social Systems 14 N°1, (2000), 81-91.

[5] J. Sjoberg, Q. Zhang, L. Ljung et al., Non linear black box modeling in system identification: An unified overview, Automatica 33 (1995) 1691-1724.

[6] P. Werbos, Beyong regression: new tools for prediction and analysis in the behavioural sciences", PhD thesis, Harvard University, (1974).

[7] D. Rumelhart, G. Hinton, R. Williams, Learning representation by back-propagating errors", Nature 323 (1986), pp. 533-536.

[8] B. M.Cottrell, and Y. Girard, M. Mangeas, C. Muller, Neural Modeling for Time Series: a Statistical Stepwise Method for Weight Elimination, IEEE Trans on Neural Networks, Vol. 6, No. 6 (1995), pp. 1355-1364.

[9] M. Cottrell, B. Girard, P. Rousset, Forecasting of curves using a Kohonen classification, Journal of Forecasting, 17: (5-6) (SEP-NOV 1998), pp. 429-439.

[10] M. Powell, Radial basis functions for multivariable interpolation: A review", J.C. Mason and M.G. Cox, eds, Algorithms for Approximation (1987), pp.143-167.

[11] T. Poggio and F. Girosi, Networks for approximation and learnin", Proceedings of IEEE 78 (1987), pp. 1481-1497.

[12] M.Verleysen and K. Hlavackova, An Optimised RBF Network for Approximation of Functions. In: Proc of European Symposium on Artificial Neural Networks (Belgium, 1994).

[13] R. Gray, Vector Quantization, IEEE Mag., Vol. 1 (April 1984), pp. 4-29.

[14] Kohonen T., Self-organising Maps, Springer Series in Information Science", Vol. 30 (Springer, Berlin, 1995).

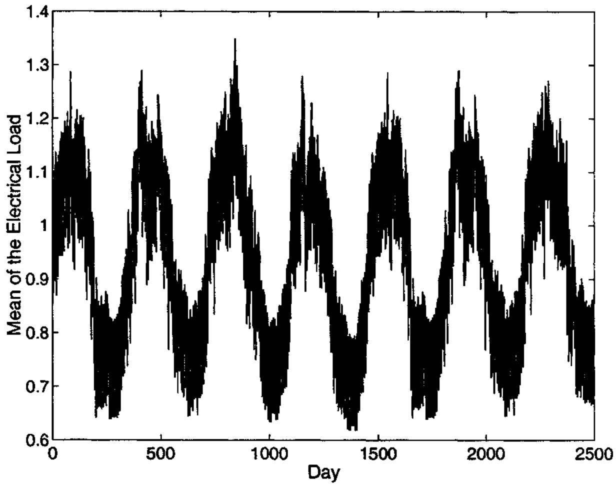

Figure 1: The mean of the Polish Electricity consumption.

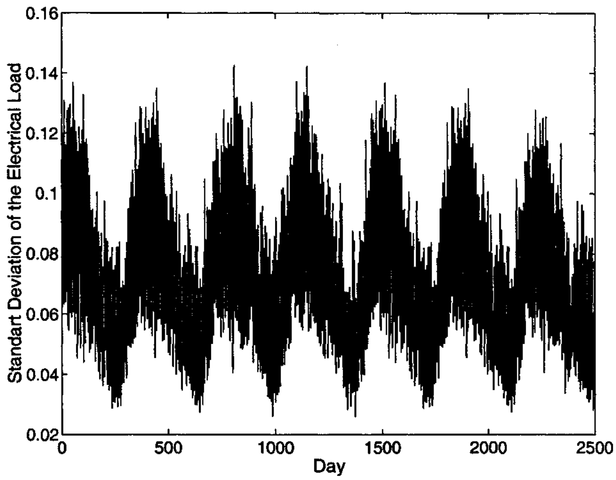

Figure 2: The standard deviation of the Polish Electricity consumption.

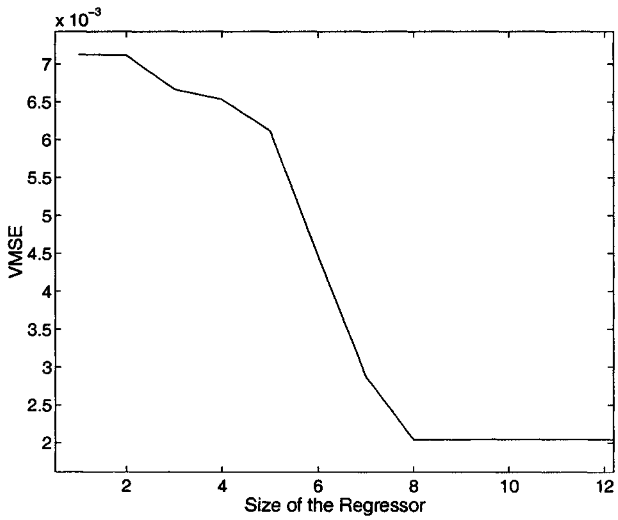

Figure 3: VMSE of a linear model for the prediction of μ_t.

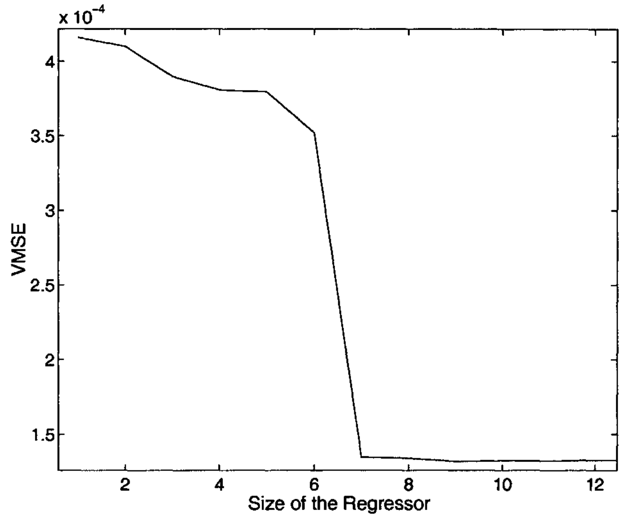

Figure 4: VMSE of a linear model for the prediction of S_t.

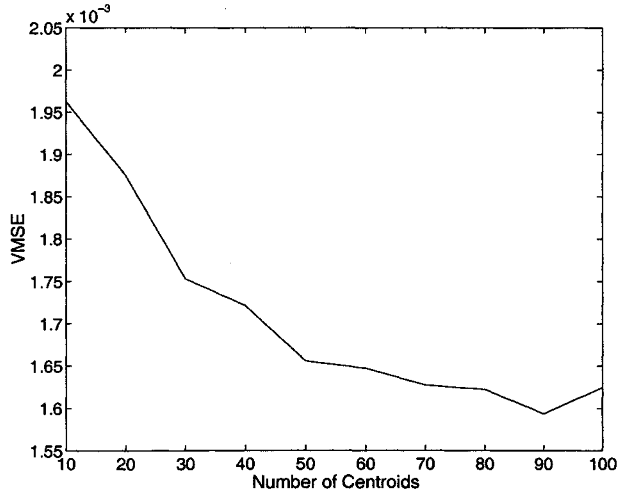

Figure 5: VMSE of a nonlinear model for the prediction of μ_t. (k fixed to 7.5)

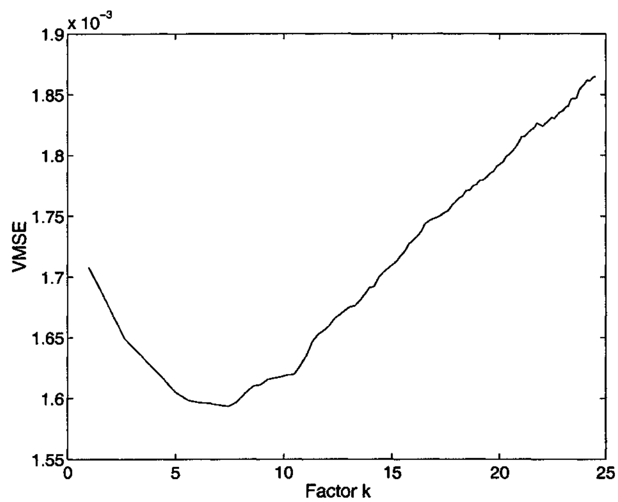

Figure 6: VMSE of a nonlinear model for the prediction of μ_t. (Number of centroids fixed to 90)

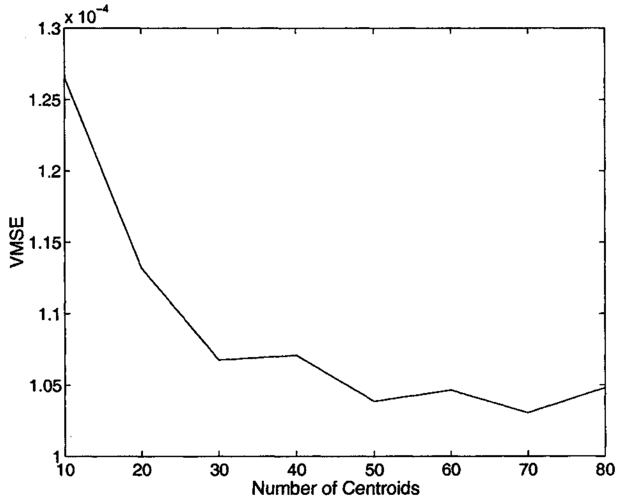

Figure 7: VMSE of a nonlinear model for the prediction of S_t. (k fixed to 2.25)

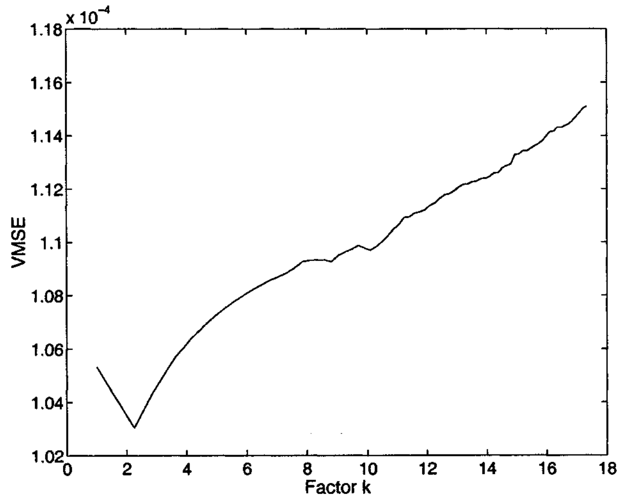

Figure 8: VMSE of a nonlinear model for the prediction of S_t. (Number of centroids fixed to 70)

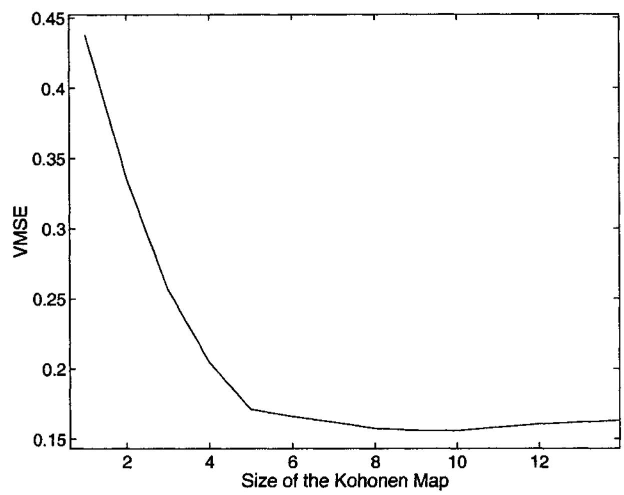

Figure 9: VMSE model for the prediction of the normalized profile.

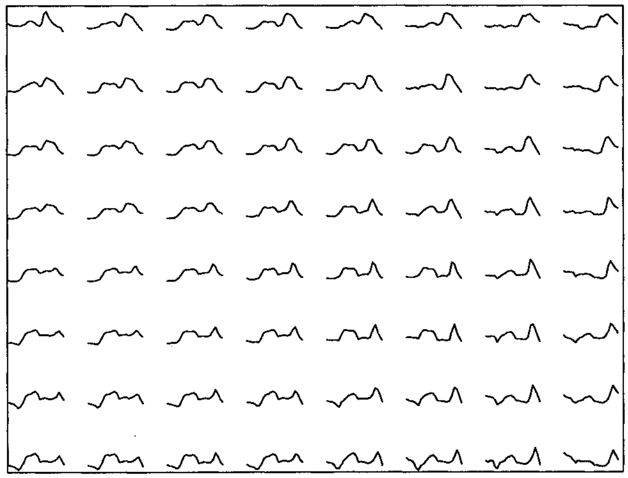

Figure 10: Kohonen map on the Polish profiles.

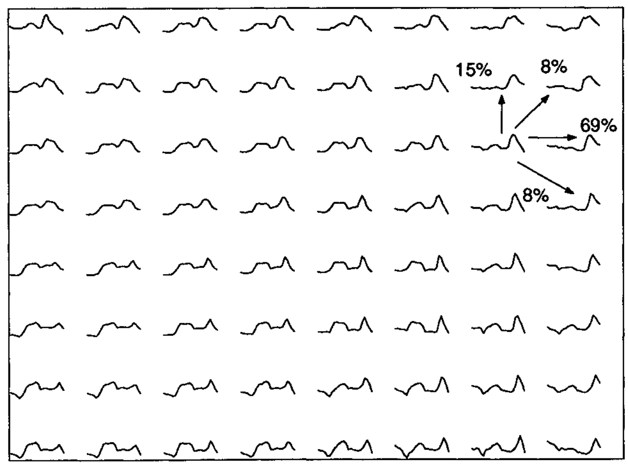

Figure 11: The probability transition between a profile and the next ones.

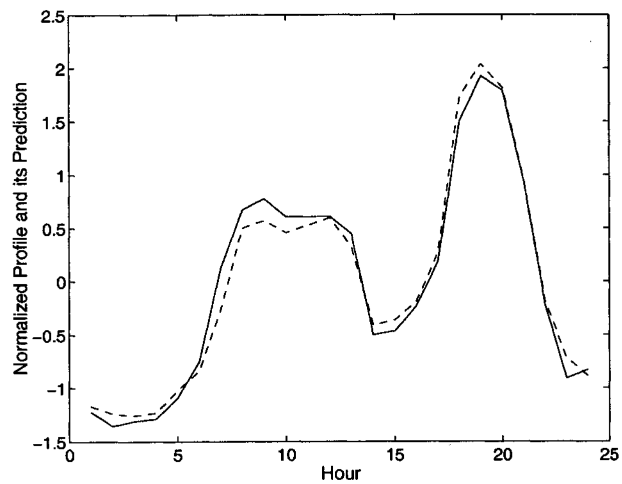

Figure 12: An Example of Normalized Profile (solid line) and its approximation (dashed line).