**World-wide competition within the EUNITE network**

**Research report**

1.Name of research people involved in the project

Peter Otto

2. Short abstract of the project

The aim of the project is it to present a method, which allows an indirect regard for actuating variables on the base of the knowledge of the time series in the past and to use it for the forecasting. In this case, one goes out from time series as seasonal dependent temporal processes.
The method allowed the forecast of future values for a time series, values of which are known from the past. The basic idea in this case consists in the fact that one can go out from that that for comparable situations in the time series the future process in this situation must occur similarly. Comparable situations in turn presuppose that also the variables actuating on the process of the time series must have been similar.
The comparision of the current situation, from which a forecast of the time series should occur, with the present measured values of the time series, will be realized over a window of determined width. The values of the current window are compared with all already measured values of the time series when the window is displaced in each case a sampling step and the differences are determined. The difference of the individual comparisions with the current window can be summerized at difference vectors.
The quality of the agreement of the respective situation on the time series with the current window is estimated for each difference vector by a distance measure, determined from a fuzzy membership function. The agreement between the current signal process and the previous signal process is expressed by the distance value D(k). The forecast value $y_{n+1}$ to be certain is determined by the $y_i$ for those D(k) is unequally zero. Consequently, the defuzzyfication occurs via the equation

$$y_{n+1} = y_n + \frac{\sum_{k=m}^{n-1} D(k) \cdot (y_{k+1} - y_k)}{\sum_{k=m}^{n-1} D(k)}$$

In this case, more similar situations from the past come in into forecast more strongly than lesser similar ones. Presupposed that the signal process is determined by the corresponding actuating variables, one can assume that the values of the actuating variables during similar signal process too must have been available.

3. Introduction to the problem- from point of view of research group

The method is full described in the next topic. For a forecast are only required the load data from 1997 and 1998. The width of the fuzzy membership function is determined automatically. The window width is determined from forecasts for the months November and December 1998.

4. The description of the approach choosen for project with details and topologies of intelligent technology used in the project

# Fuzzy based time series forecasting of electric load

**P. Otto**
Technische Universität Ilmenau
Postfach 10 0565, 98684 Ilmenau
Telefon: + 49 (0) 3677/69-27 73
Telefax: + 49 (0) 69-14 34
e-mail: otto@systemtechnik.tu-ilmenau.de

**Keywords:** forecast, time series, fuzzy-methods, electric load

## Abstract

The paper presents a new method for the forecast of time series. This method requires no model of the signal process. It is based on an arrangement of the current situation with situations from the past. With the aid of a fuzzy-processing algorithm, the forecast values are determined from this arrangement. The effectivness is shown at an example for electric load forecast of a power distribution company.

## 1 Introduction

The prediction of time series is an very important problem in monitoring, diagnosis, control and decision support for technical and nontechnical systems [ander], [young], [andel]. In addition to specific time-dependent (seasonal) actuating variables for example temperature, global solar radiation, daily flow of life and so on, such signals are often subject of not or only hardly registrable actuating variables so that a prediction is combined with great insecurities. The aim of this paper is it therefore to present a method, which allows an indirect regard for actuating variables on the base of the knowledge of the time series in the past and to use it for the forecasting. In this case, one goes out from stationary time series as seasonal dependent temporal processes. The method based on the application of a fuzzy-algorithm and will be demonstrated on an example of the load forecasting in a power distribution company. The fuzzy-prediction algorithm is realised in an software tool "FuzzyPredict", written in The method allowed the forecast of future values $y_{n+1}$, $y_{n+2}$, ... for a time series $\{y_1, y_2, ... y_n\}$, values $y_i$ ($i \leq n$) of which are known from the past. The basic idea in this case consists in the fact that one can go out from that that for comparable situations in the time series the future process in this situations must occur similarly. Comparable situations in turn presuppose that also the variables actuating on the process of the time series (for example temperature, global solar radiation and others) must have been similar. In addition, one can conclude besides from this that also non measurable actuating variables (for example the consumer behaviour in the case of consumption of electric energy from a power distribution company) must have been similar. In such a way non measurable actuating variables are considered implicit in comparable situations. The comparison of the current situation at the time n @ T (T - sampling time), from which a forecast of the time series should occur, with the present measured values of the time series, will be realised over a window of determined width m. Depending on dynamics (memory length) of the signal process, the width of the window can be chosen differently. The values of the current window ($y_{n-m+1}$, $y_{n-m+2}$, ..., $y_n$) are compared with all already measured values of the time series when the window is displaced in each case a sampling step and the differences $e_{k,i} = y_{n-k} - y_{n-i+1}$ ($i = 1, ..., m$; $k = m, ... n - 1$) are determined (figure 1).

The difference of the individual comparisons with the current window can be summarized at difference vectors $\underline{e}_{n-1}$, $\underline{e}_{n-2}$, ... $\underline{e}_m$. The quality of the agreement of the respective situation on the time series with the current window is estimated for each difference vector by a distance measure, determined from a fuzzy membership function (figure 2) [bockl].

# 2 The Fuzzy-Forecasting Method

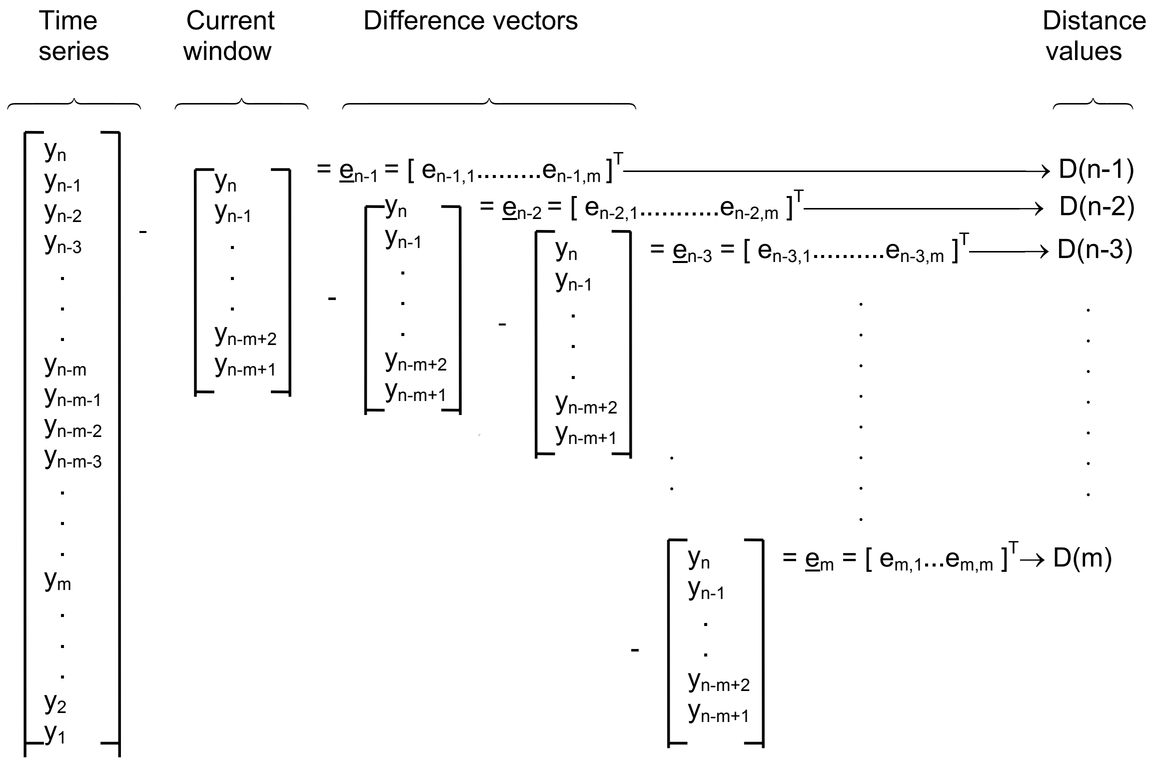


Figure 1: Formation of the difference vectors $\underline{e}_k$

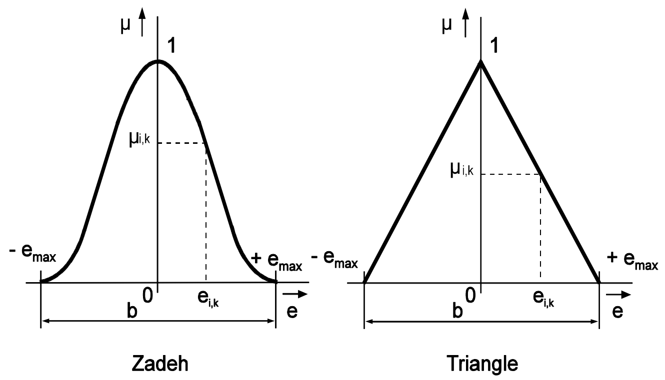


Figure 2: Fuzzy membership functions for the appraisal of the differences $e_{k,\,i}$

The width of the membership function (2 @ $e_{max}$) is determined from the demands at the forecast quality. The more precisely the forecast should occur, the width (2 @ $e_{max}$) must be chosen the less. The agreement between the current signal process (at the actual time n @ T) and the previous signal process (at the time i @ T, with i = n – 1, ..., m; m < n) is expressed by the distance value (figure 1):

$$D(k) = \prod_{i=1}^{m} \mu_{k,i} \qquad (k = m \ldots n-1) \qquad (1)$$

In the case of full agreement of all values of the current window with the situation of the past, D(k) accept the value 1.
Is only one difference $e_{k,\,i}$ outside of maximum width of the membership function μ, D (k) accept the value 0. The more D(k) approaches the value 1, the agreement between current signal process and the considered situation from the past is all the more greater. The distance values D(k) determined in such a way are now used, via an inference mechanism of the form IF ... THEN, to determine the next values of the time series to be predicted:

***IF*** *the actual signal process is comparable to occurred signal process in the past at time i @ T*

***THEN*** *the next value of the time series $y_{n+1}$ is chosen similar to the value $y_{i+1}$*

*Or more detailed:*

$$\begin{array}{ll} \textbf{\textit{IF}} & y_n \approx y_i \ AND \ \ y_{n-1} \approx y_{i-1} \ \ AND \ \ y_{n-2} \approx y_{i-2} \ldots \\ & AND \qquad y_{n-m} \approx y_{i-m} \\ \textbf{\textit{THEN}} & y_{n+1} \approx y_{i+1} \end{array} \quad (2)$$

*With $i = n-1, n-2, \ldots, m; m < n$.*

A measure of the applicability of this IF ... THEN-rule is D(k). The forecast value $y_{n+1}$ to be certain is determined by the $y_i$ for those D(k) is unequally zero. Consequently, the defuzzyfication occurs via the equation:

$$y_{n+1} = \frac{\sum_{k=m}^{n-1} D(k) \cdot y_{k+1}}{\sum_{k=m}^{n-1} D(k)} . \quad (3)$$

In this case, more similar situations from the past come in into forecast more strongly than lesser similar ones. Presupposed that the signal process is determined by the corresponding actuating variables, one can assume that values of the actuating variables during similar signal process too must have been available. In such a way, it is possible to consider the effect of these actuating variables on the signal process in the case of forecast without registering them directly. That of course also applies to the not measurable actuating variables.

# 3 Application of the Fuzzy-Time Series Forecast for the Load Forecast

The electricity consumption profile of a power distribution company can after elimination of possible long-term trends regarded as a stationary time series with seasonal characteristic. For these companies it is extremely important for an optimal management of the purchase and of the sale of electric energy to have an as precise as possible short-time prediction of the load to be expected [rausch]. However, it is opposed that this process, in addition to the external actuating variables (for example temperature, global solar radiation, wind force and others), is embossed by the habits and orders of the consumers (both industrial and private) very strongly. In such a way, the loadprocess distinguishes from weekday at weekday, at weekends, at vacation times ore on holidays and so on. Therefore, a load forecast with traditional models (ARIMA, ARMAX or others) [box], [mast], in particular on such "not usual" days, is combined often with great mistakes or insecurities. For this reason, one manages often that one employs several models or includes "model days" which were determined by a classifier in forecast [bret]. The possibility of the load forecast should be represented here at an example by means of Fuzzy-time series forecast. A time series which contains load values from a period of two years (1996, 1997)were employed to this.
The sample time conducted T = 1 hour, so that 24 one hour values result per day for the load. m = 6 was chosen as width of the window. The basic width of the membership function was $2 * e_{max} = 100$.
The results of forecast are represented in figures 3-6.
Figure 3 shows a one hour - forecast of the electric load for the period dated March 27th. by April 17th, 1996.
This period contains "normal" weekdays and "normal" weekends as well as holidays (Good Friday, Easter Sunday and Easter Monday) and vacations days both (1.4. - 13.4).
In addition, the changeover from the winter time occurred on summer time in the night from 30.3 by 31.3.
The basis for forecast was the period from 1.1. by 16.4.1996.
One can recognize that forecast supplies the best results on "normal" weekdays (standard deviation of the prediction error $S_e = 41{,}56$) while she is worse at "normal" weekends (Se = 46,61 ).
That is to be traced back to the fact that considerably more model days are available for the forecast on weekdays as for weekends. Forecast is proved even worse on the holidays because no model days are available for this in the employed data ($S_e = 61{,}04$).
Forecast yet provides completely preposterous values since an averaging of similar situations occurs nevertheless from the past.
Figure 4 shows the forecast errors for the considered period.
The one hour - forecast for a comparable period from 1997 is represented in figure 5 (19.3. - 9.4.).
In this case, the time series for the electric load dated 1.1.1996 by 18.3.1997 were employed for forecast.
Although it is to be seen too here that the best results are achieved on "normal" weekdays (Se = 26,67) and the worst ones are achieved on holidays (Se = 51,99), an improvement can always be found compared to the forecast from 1996 .
That is to be traced back to the fact, that a considerably greater number of model days from the past can be employed.
Even with available measurement errors, as for example with time changeover (night of the 29th by 30.4.1997), reasonable forecasts are made. The width of the membership function in this case was extended however to $e_{max} = 800$ in order to find suitable model days for forecast from the past.
After this was the width of the membership function again on $e_{max} = 100$ reduced.
The forecast errors for the considered period are represented in figure 6.

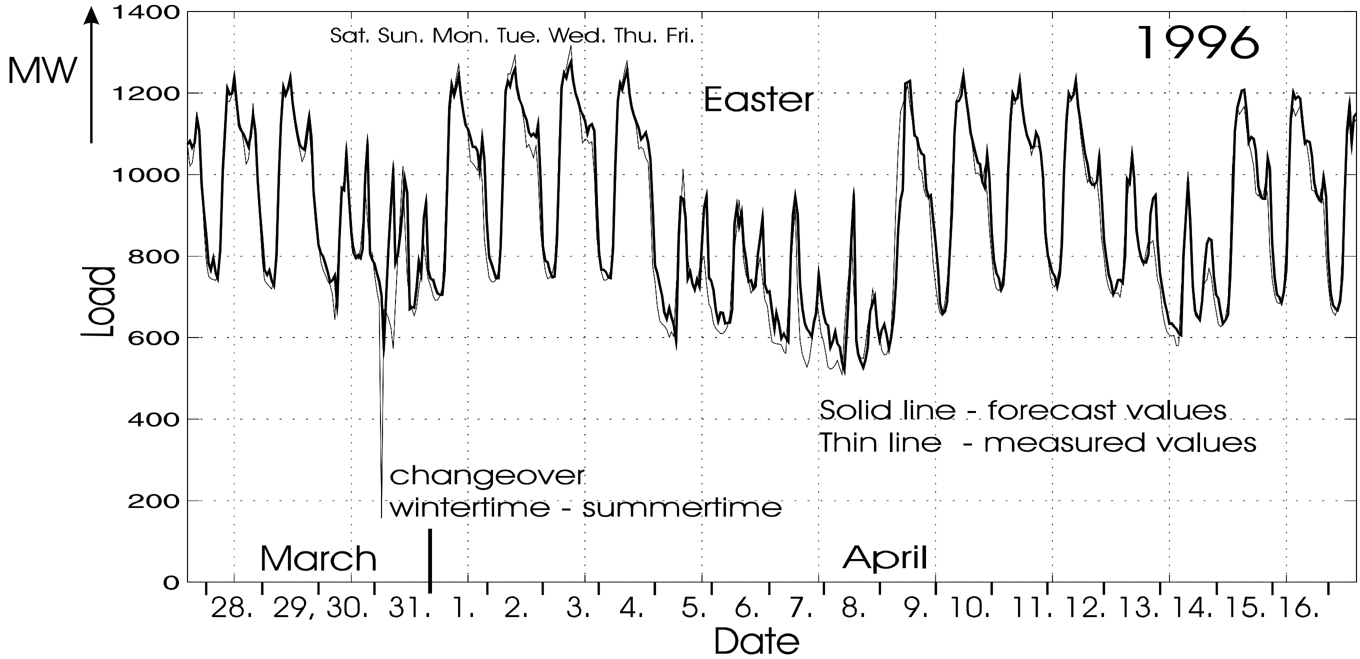


Figure 3: One sampling step forecast of the load for three weeks in 1996

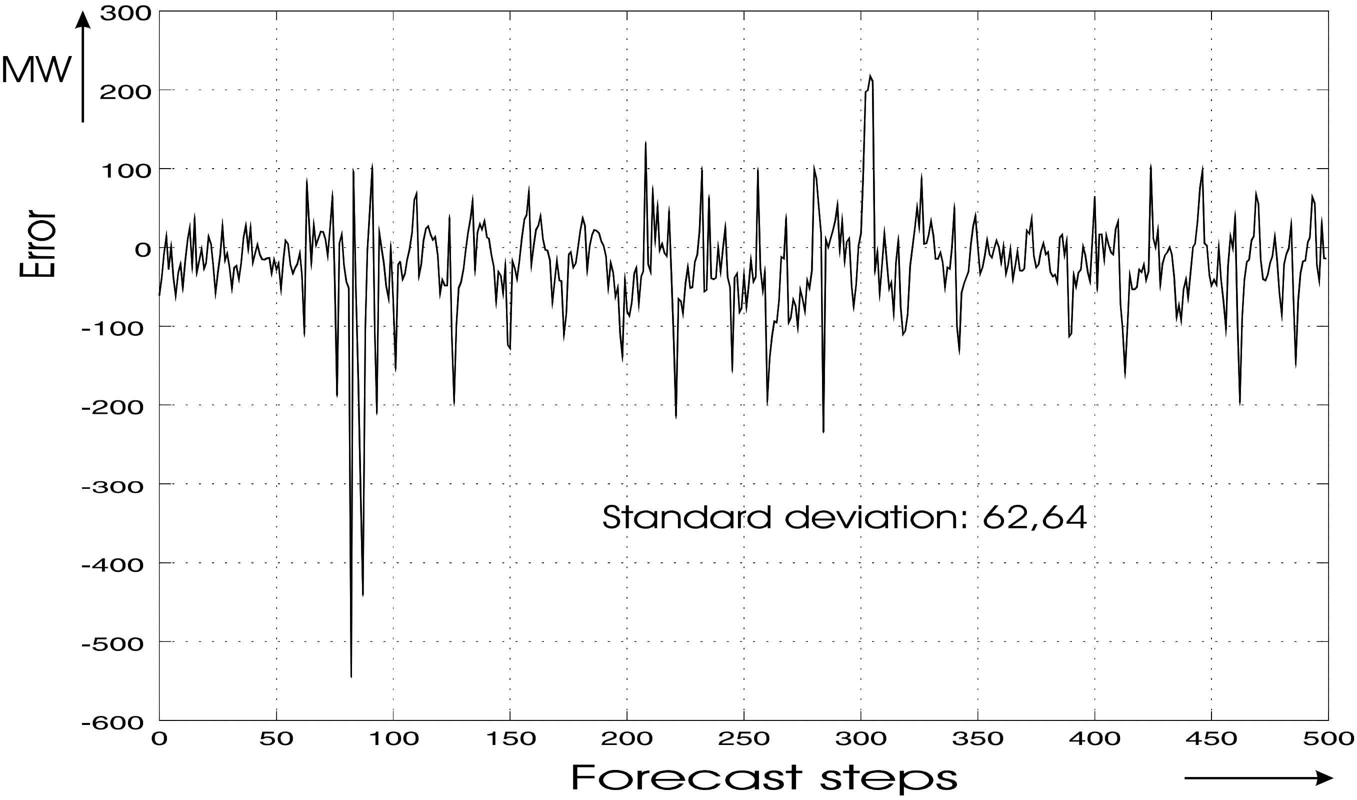


Figure 4: Forecast error of the load for three weeks in 1996

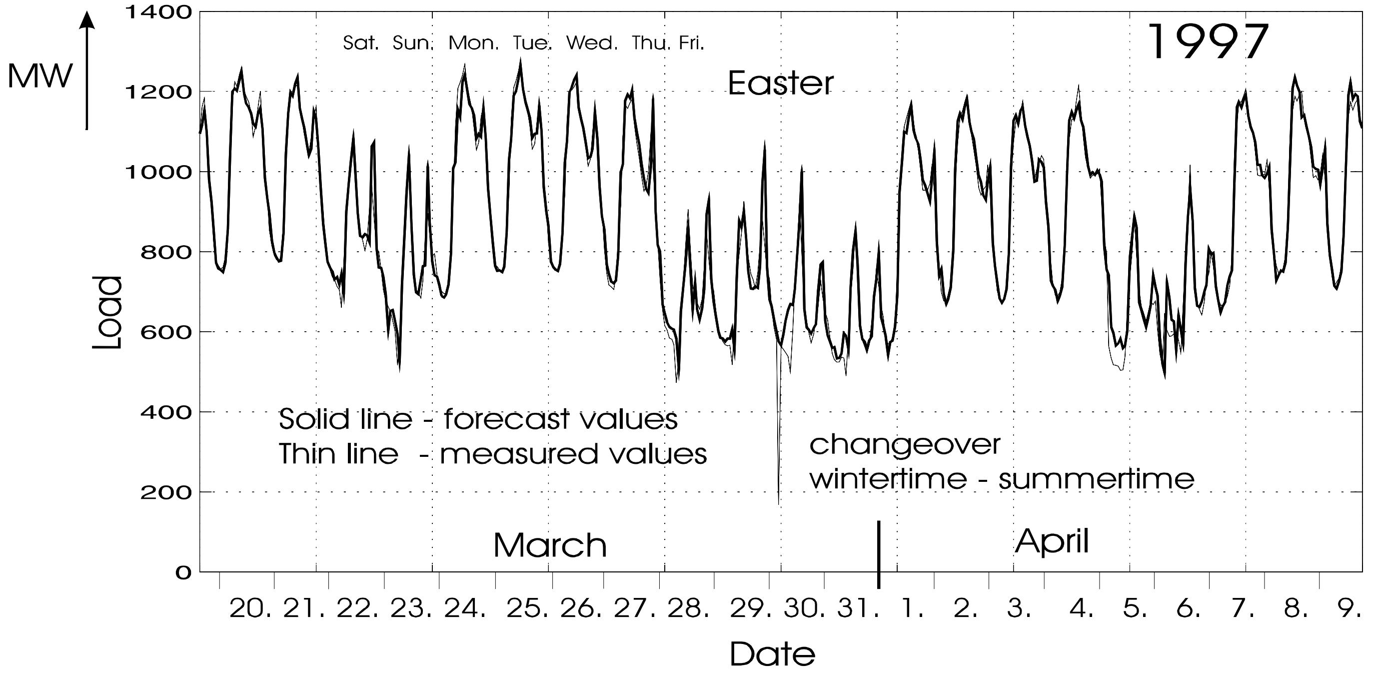


Figure 5: One sampling step forecast of the load for three weeks in 1997

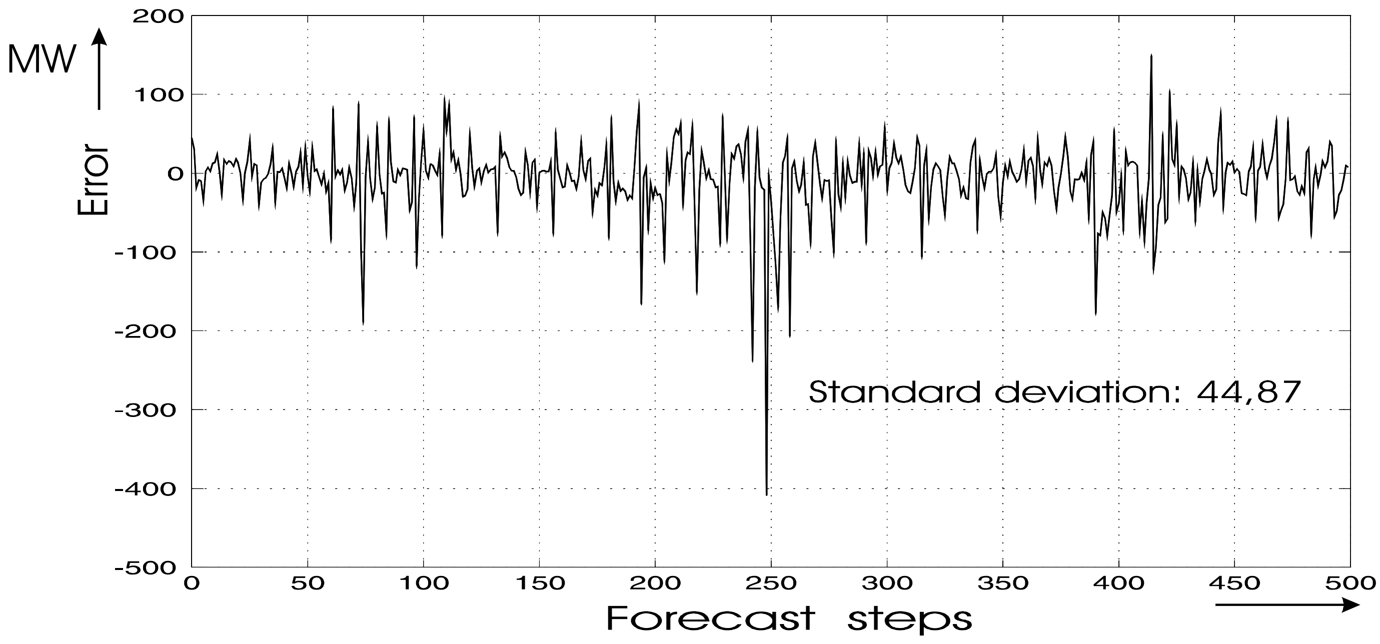


Figure 6: Forecast error of the load for three weeks in 1997

# 4 Conclusion

The previous investigations showed that the Fuzzy-time series forecast supplies good results. That could be shown at examples of the load forecast of three weeks from 1996 and 1997 of a power distribution company. The main advantage of the method exists in it that one completely searches autonomous for agreements of the current situation (of that a forecast should occur) with the past and from it the forecasting values determined may be. Therefor, a previous classification in different model days is not necessary so that a system of rules can also be dropped for the inclusion of these model days in forecast. The otherwise necessary inclusion of not measurable actuating variables in the form of these model days, is here replaced by its indirect regard via similarity of the signal processes. Assumption is natural that a sufficiently big history of the time series is available and the width of the membership function is not chosen too small so that all computed distance values D(k) do not become immediately zero. The described Fuzzy-time series forecast algorithm was implemented in a software tool "FuzzyPredict", written in C++. For an optimization of forecast, further investigations are still necessary with regard to the optimal choice of the window width m and the width of the membership function 2 @ $e_{max}$

# 5 References


[ander] Anderson, O. D., "Forecasting", North-Holland Company, London Amsterdam, (1979)

[young] Young, P. C., "Recursive Estimation and Time Series Analysis", Springer Verlag Berlin,(1984)

[andel] Andel, J., "Statistische Analyse von Zeitreihen", Akademie Verlag Berlin, (1984)

[bockl] Bocklisch, S. F., "Prozeßanalyse mit unscharfen Verfahren", Verlag Technik Berlin, (1987)

[rausch] Rauschenbach, Th., Wernstedt, J., Hartig, G., Palm, M., "Anwendung von wissensbasierten Methode zur Lastprognose und zum Lastmanagement in einem EVU", VDI-Tagung "Informationserfassung und –verarbeitung in der Energietechnik", *VDI-Berichte* Nr. 1210**,** Dresden, (1995)

[box] Box, G., Jenkins, G., "Time-series Analysis, Forecasting and Control, Prentice Hall Englewood Cliffs NJ., (1976)

[mast] Masters, T., " Neural, Novel & Hybrid Algorithms for Time Series Prediction", John Wiley & Sons Inc., (1995)

[bret] Bretschneider, P., Rauschenbach, Th., Wernstedt, J. "Hybrid Forecast strategy using an Adaptive Fuzz Classification Algorithm", *EUFIT '98,* Proceedings Volume 3, pp. 1916-1919, (1998)

5. The description of the data and its handling as training and testing data

A training phase is not necessary in the case of the described method. The adaptation of the window width occurred with tests for the months November and December 1998.

6. Experimental results on the verification projects using provided data

In the case of the tests, the values for all time steps (not only the daily maximum values of electrical load) lay in the following fields:

MAPE: 4,41 % - 7,69 % (n = 1488)

M: 150 – 265 (i = 1,......., 1488)

The forecast for January 1999 was carried out with time steps of a half hour with a window width of 250 and a width of the fuzzy membership function of 60. The maximal value for every day was determined from it. The results are contained in the file “Jan99.dat” (ASCII-file).

7. Summary of the project with comments about features of technologies used in the approach.

See section 4.