Tr. J. of Engineering and Environmental Science
23 (1999) , 173 – 180.

A New Fuzzy Modelling Approach For Predicting The Maximum Daily Temperature From A Time Series

Hasan TATLI, Zekai ŞEN
Istanbul Technical University, Meteorology Department,
Maslak 80626, Istanbul-TURKEY

Received 17.09.1998

Abstract

Classical time series analysis requires many assumptions such as the normality of data, linearity in the autocorelation coefficient and statistical parameter estimations. It is almost impossible to find all these assumptions applicable in stochastic time series generation or simulation. This paper provides a simple fuzzy-probabilistic method for the time series analysis. The basis of the methodology is to construct the fuzzy base rule domain from the available daily maximum temperature records at Kandilli observatory in Istanbul. The new concepts of transition and cumulative probability procedures are employed for taking decision among the alternative consequent fuzzy sets prior to the defuzzification.

Key Words: Fuzzy rule-base, modeling, probabilistic, time series, transition matrix.

Günlük En Büyük Sıcaklığın Zaman Serisinden Yeni Bir Bulanık Modelleme Yaklaşımı İle Tahmini

Özet

Klasik zaman serilerinin incelenmesi verinin normalliği, öz ilişki katsayılarının doğrusallığı ve istatistiksel değişkenlerin kestirimi gibi kabuller gerektirir. Stotastik zaman serilerinin oluşturulması veya simulasyonunda, bütün bu kabullerin uygulanabilirliği hemen hemen imkansız gibidir. Bu çalışma, zaman serilerinin incelenmesinde, basit bir bulanık-olasılık yöntemi sağlamaktadır. Metodoloji, İstanbul Kandilli rasathanesinde kaydedilen günlük en büyük sıcaklıklardan, bulanık kural temeli oluşturulmasına dayanmaktadır. Durulaştırma adımından önceki çeşitli bulanık küme seçeneklerine karar vermek amacı ile yeni yapılı geçiş ve birikimli olasılık prosedürü uygulanmıştır.

Anahtar Sözcükler: Bulanık kural, geçiş matrisi, olasılık, zaman serisi

Introduction

The pioneering work of Zadeh (1965) concerning the processing of the linguistic uncertainties by the of fuzzy sets has opened a wide spectrum of applications in many diverse fields. Fuzzy application areas include estimation, prediction, control, approximate reasoning, intelligent system design, machine learning, image processing, machine vision, pattern recognition, medical computing, robotics, optimization, civil, chemical and industrial engineering. Unfortunately, fuzzy applications in meteorology domain are rather very rare and there is a great future in its application to atmospheric and meteorological problem solutions. The atmospheric events are complex, ambiguous and vagueness embedded in their nature. This is mainly due to the fact that earth and atmospheric scientists are involved basically with traditional uncertainty techniques among which are the statistics, probability and stochastic processes with control implementations through adaptive Kalman filtering. However, there is an unlimited scope application possibilities in natural sciences for the fuzzy principles.

Fuzzy techniques for treating uncertain qualitative information include fuzzy set theory, fuzzy arithmetic and mathematics, fuzzy logic, fuzzy decision making and fuzzy control. In general fuzzy procedures transform through uncertain basic rules that reflect the behavior of the system concerned and consequently the uncertain or crisp information as initial and boundary conditions as well as the input variables are mapped so as to produce again uncertain or crisp results. In any natural even very precise approaches that we think are accurate include a certain amount of ambiguity and vagueness. Suggestion of the chaos theory for dynamic systems where the system equations are the fundamental laws of the physics and the conservation principles of energy, momentum and mass in addition to the thermodynamic fundamentals the solutions are always dependent on the initial conditions. Although the system remain the same, infinitesimally small changes in the initial conditions lead to different results, which can be regarded collectively as numerically vague and ambiguous. This is tantamount to saying that even though the system equations are spatially and temporally the same, they transfer negligibly close input values to completely independent and very different output values. Hence, the question is what causes such randomness, unpredictability in the output values? The answer is the vagueness in the system equations which cannot reflect the reality exactly or which cannot adopt itself to the small changes during the evolution of the natural phenomenon. It is therefore logical to treat these phenomena by the fuzzy principles.

(i) Another elegance of the fuzzy set theory is that during the assimilation of input data it does not require any specification concerning the data structure. For instance in the statistical or stochastic modeling if the data is distributed according to the normal (Gaussian) distribution then the available stochastic procedures in the data treatment can be used. Otherwise, prior anything the data must be rendered into a normal form.

(ii) In fuzzy treatment linguistic rules are utilized to approximate the desired output or predictions.

(iii) The construction of model does not require any integro-differential equations or recurrence relationships similar to the Markov or ARIMA models.

Jenkins and Watts (1968) give detailed account for time series analysis in time domain by Box and Jenkins (1970) and in the frequency domain. Aforementioned drawbacks in clissical time series analysis are not encountered in the fuzzy probabilistic model proposed in this paper. Besides, the classical techniques require separation of trend and periodicity prior to stochastic prediction in the time series. Fuzzy probabilistic method, on the other hand, does not require such separations.

This paper shows the application of fuzzy logic to daily maximum temperature sequences recorded at Kandilli observatory in Istanbul.

Fuzzy Sets

Sets are collection of objects with the same properties and in crisp sets the objects either belong to the set or not. In practice the characteristic value for an object belonging to the set considered is coded as 1 and if it is outside the set then the coding is 0. For instance, in a set of positive even integer numbers the first three objects are 2, 4 and 6 their schematic representations with characteristic values are shown in Figure 1a.

In crisp sets, there is no ambiguity or vagueness as for the belonging of each object to the set concerned. On the other hand, in daily life human are always confronted with objects that may be similar to each other with different properties and therefore there arises uncertainty as to their belonging to a common set with membership values 0 or 1. Of course, logically some of the similar objects may partially belong to the same set and therefore, an ambiguity emerges in the decision of belonging or not. In order to alleviate such situations Zadeh (1965) generalized the crisp set membership degree as having any value continuously between 0 and 1. Fuzzy sets are a generelization of conventional set theory. The basic idea of fuzzy sets is easy to grasp. Hence an object with membership degree 1 belongs to the set with no doubt and those with 0 membership values again absolutely do not belong to the set but objects with intermediate membership degrees belong to the same set partially. The greater is the membership degree the more the object belongs to the set. For instance, if approximately positive even integer numbers are requested then the membership function in Figure 1a takes the form in Figure 1b where there are interference between the numbers desired because of a fuzzy linguistic word **approximately**. In this manner any fuzzy linguistic word can be expressed as fuzzy set. In meteorology there are many linguistic fuzzy words some of which are **warm, cloudy, foggy, dense, high, low, dry, wet, small, etc.**

In meteorology, for instance, any statement about the weather temperature includes uncertainty in the forms of vagueness or ambiguity. If the temperature at a place changes between almost T_0 and T_1 °C then this domain of change should have linguistically some subsets by considering everyday conservation. In general the temperature is either **cold** or **cool** or **warm** or **hot**. Hence, there are four subsets of the temperature universal set at a location. Within the whole universal set it is not possible to define the delimitation of these linguistic words with certainty. However, intuitively one can know approximate position of each linguistic word as a shown in Figure 2.

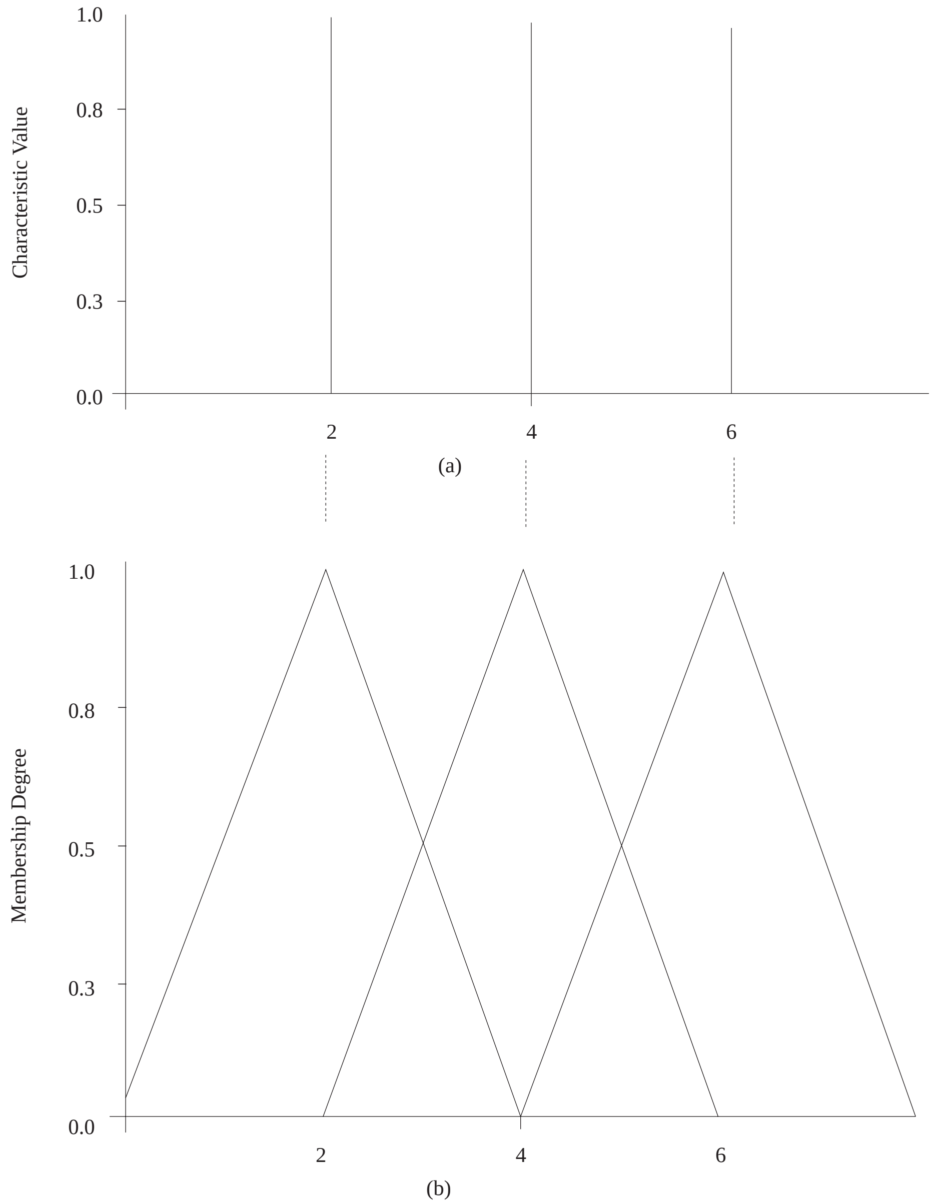

Figure 1. Crisp and fuzzy numbers

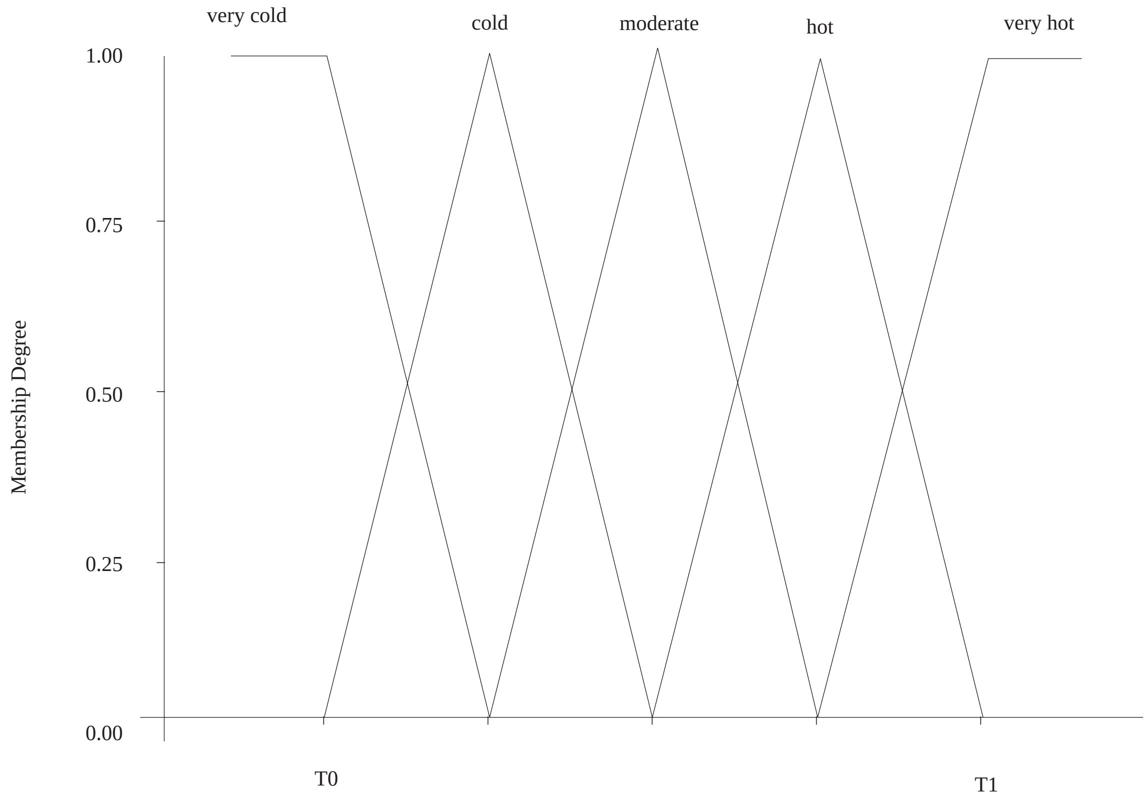

Figure 2. Temperature fuzzy subsets

Accordingly constructed triangles represent the approximate properties of cold, cool, warm and hot fuzzy subsets. Any meteorological factor can be subdivided into fuzzy sets that interfere with each other. However, a subjective point in delimiting the fuzzy subsets can be avoided by employing actual data and/or expert opinions as will be explained in the application section of this paper.

1. Fuzzy Rule Base

In any diagnostic or prognostic study in meteorology for the application of fuzzy reasoning there are three interdependent steps. A successful execution of these steps leads to the solution of the problem in a fuzzy environment, i.e., the solution procedure digests any type of uncertainty in the basic evolution of the event concerned.

(a) Fuzzification step: All meteorological events are considered as having ambiguous characteristics and therefore their domain of change are divided into many fuzzy subsets which complete, normal and consistent with each other. Hence the domain of change is fuzzified. This stem is applied to each meteorology factor considered in the solution of the problem.

(b) Inference: This step is, in fact, relates systematically pair wise all the factors that take place in the solution depending on the purpose of the problem. In fact this part includes many fuzzy conditional statements to describe a certain situtation. For instance if two events **X** and **Y** are interactive then they are dependent on each other. Conditional statements express the dependence as follows verbally without any equation as used in the classical approaches,

$$IF\ \ X\ \ is\ \ A\ (1)\ THEN\ \ Y\ is\ B(1)$$

$$ALSO$$

$$IF\ \ X\ \ is\ \ A\ (2)\ THEN\ \ Y\ \ is\ \ B\ \ (2)$$

$$ALSO$$

$$IF\ X\ is\ A\ (3)\ THEN\ Y\ is\ B(2) \tag{1}$$

$$ALSO$$

$$\cdots$$

$$\cdots$$

$$\cdots$$

$$ALSO$$

$$IF\ X\ is\ A\ (n)\ THEN\ Y\ is\ B\ (n)$$

where A(1), A(2),. . ., A(n) and B(1), B(2),. . ., B(n) are the linguistic description of X and Y respectively, and they are fuzzy subsets of X and Y that cover the whole domain of change of X and Y. The fuzzy conditional statements in Eq.(1) can be formalized in the form of the fuzzy relation R(X,Y) as R(X,Y)=ALSO $(R_1, R_2, R_3, \ldots, R_N)$ where ALSO represents a sentence connective which combines R_i's into the fuzzy relation R(X,Y), and R_i denotes the fuzzy relation between X and Y determined by the i-th fuzzy conditional statement. After having established the fuzzy relationship R (X,Y) then the compositional rule of inference is applied to infer the fuzzy subset B for Y, given a fuzzy subset A for X as B=AoR(X,Y) where "o" is a compositional operator (Kosko, 1992).

(c) Defuzzification: The final result from the previous step is in the form of fuzzy statement and in order to calculate the deterministic value of a linguistic variable Y the defuzzification method must be applied (Kiszka, et al., 1985a,b) as

$$y = \frac{\sum_i^L y_i}{L} \tag{2}$$

or center-average method according to Wang (1993) via using fuzzy basis expansion, expressed as

$$p(x) = \frac{\prod_{i=1}^M \mu_{A_{ij}}(x_i)}{\sum_{j=1}^L \prod_{i=1}^M \mu_{A_{ij}}(x_i)} \tag{3}$$

$$y = f(x) = \sum_{j=1}^L p_j(x) y_j \tag{4}$$

where p(x) is fuzzy basis function and y is particular value of the linguistic variable Y, y_j is the support value in which the membership function reaches its maximum grade of membership, and finally L is number of rules and M the number of inputs.

2. Application

Fuzzy controller by means of the first order structural dependence along a given time series provides simple prediction process provided that the fuzzy subsets of the variability domain is divided into meaningful fuzzy intervals. The application of the methodology proposed in the previous sections is presented for daily temperature records at Kandilli observatory where records are kept since 1912. Such a long record is not necessary for the model development hence, in this study, only the most recent two years duration daily maximum temperature records are used. First of all, the maximum temperature domain is divided into 11 triangular subsets that are normal, consistent and complementary. Here, normality implies that fuzzy subset has membership value equal to 1 at least for one of the members. They are complementary in the sense that at any temperature value there are distinctive fuzzy temperature subsets and their membership degress summation at a given temperature is equal to 1. On the other hand, these 11 fuzzy subsets, namely A_i (i=1,2,. . . ,11) are shown in Figure 3 and they are treated equivalently for the input and output maximum temperature values. Herein, the input temperature is the maximum temperature of any day and the output the maximum temperature for the following day.

Once the fuzzy temperature subsets are provided, it is then possible to train sequentiall the temperature time series so as to find the steady state percentages, i.e., probabilities in the transitional matrix. The first 365 daily temperature values are employed for determining the transition matrix elements from the fuzzy subsets in Table 1. These are final fuzzy associative matrix elements.

After training period the following fuzzy rule base is obtained for the maximum temperature prediction.

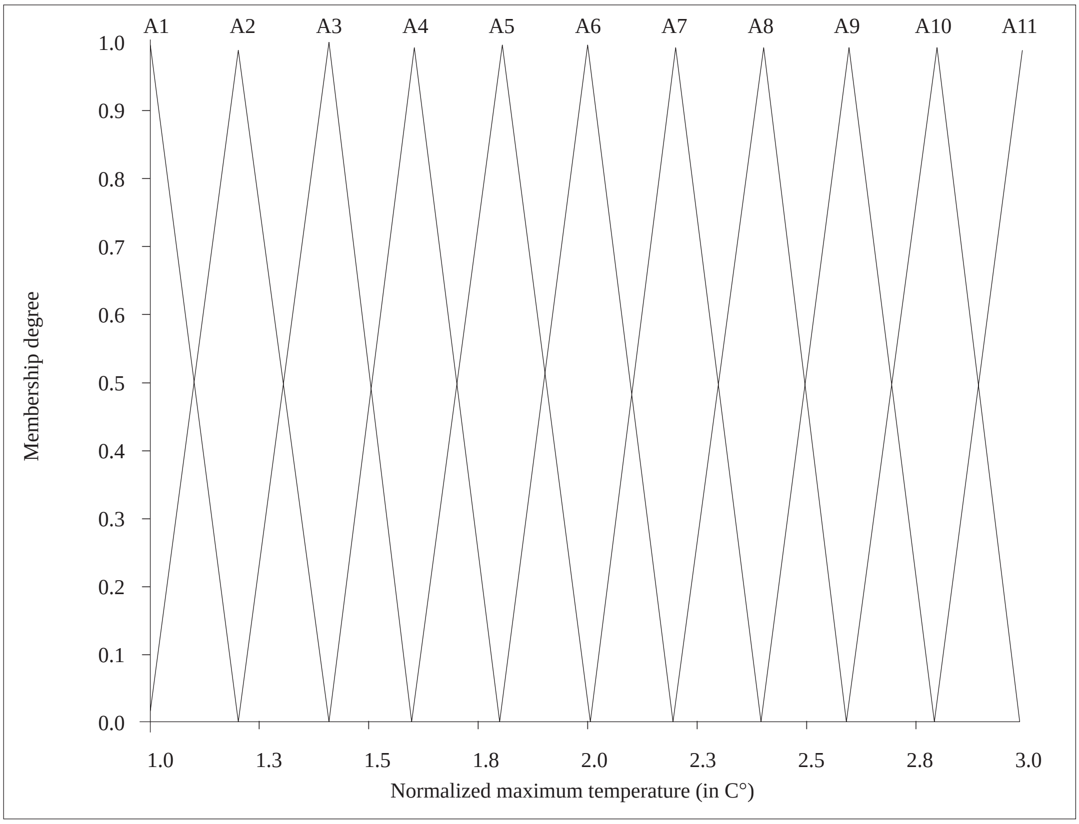

Figure 3. Kandilli maximum temperature fuzzy subsets

Table 1. Relative transition matrix

	A1	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**A8**	**A9**	**A10**	**A11**
A1	0.3	0.5	0.2	0	0	0	0	0	0	0	0
A2	0.06	0.29	0.43	0.21	0.01	0	0	0	0	0	0
A3	0.01	0.13	0.36	0.35	0.13	0.02	0.01	0	0	0	0
A4	0	0.04	0.23	0.37	0.24	0.08	0.03	0.00	0	0	0
A5	0	0	0.10	0.28	0.34	0.21	0.06	0.01	0.00	0	0
A6	0	0	0.01	0.09	0.22	0.034	0.23	0.07	0.03	0.01	0.00
A7	0	0	0	0.01	0.07	0.29	0.36	0.18	0.06	0.02	0.01
A8	0	0	0	0	0.00	0.07	0.22	0.37	0.27	0.06	0.01
A9	0	0	0	0	0	0.01	0.07	0.32	0.42	0.18	0.01
A10	0	0	0	0	0	0	0.01	0.15	0.46	0.34	0.03
A11	0	0	0	0	0	0	0	0.36	0.5	0.14	

This fuzzy rule-base is the main tool in prediction the future likely maximum temperature values. Figure 4 gives the trained and predicted temperature sequences. It is obvious that the periodic pattern in the daily temperature sequences is modeled succesfully with the proposed fuzzy logic model, because the relative error appears less than 10 percent. In the non-training part, the actual values and predicted ones do not fall on each other and consequetly the error amount is X-$\hat{X}$. However, in order to asses the validity of fuzzy prediction it is necessary to have an overall measure of the individual errors in the form of average performance error (APE) defined as follows

$$APE = \frac{\sum_i^n |x_i - \hat{x}_i|}{\sum_i^n |x_i|} \times 100 \tag{5}$$

For the daily temperature series calculated in this study APE=% 7.12. This is less than practically acceptable limit of 10%.

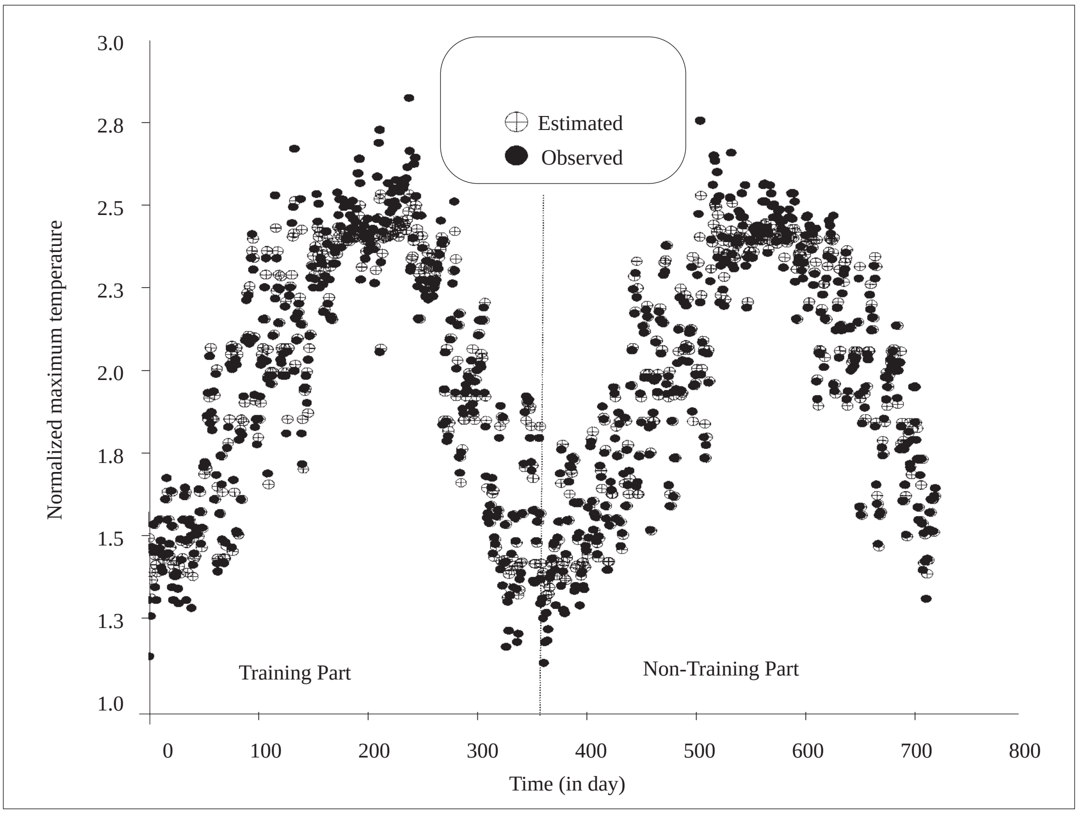

Figure 4. Training and prediction parts of maximum temperature time series

Table 2. Fuzzy rule-base

IF	X(t)	A1	**THEN**	X(t+1)	A2
IF	X(t)	A2	**THEN**	X(t+1)	A3
IF	X(t)	A3	**THEN**	X(t+1)	A3
IF	X(t)	A4	**THEN**	X(t+1)	A4
IF	X(t)	A5	**THEN**	X(t+1)	A5
IF	X(t)	A6	**THEN**	X(t+1)	A6
IF	X(t)	A7	**THEN**	X(t+1)	A7
IF	X(t)	A8	**THEN**	X(t+1)	A8
IF	X(t)	A9	**THEN**	X(t+1)	A9
IF	X(t)	A10	**THEN**	X(t+1)	A10
IF	X(t)	A11	**THEN**	X(t+1)	A11

3. Conclusions

Prediction of meteorological records such as the maximum temperature serially from the structure of the observed time series the fuzzy rule-base modelling provides an efficient way. This approach does not give any recurrence formulation like Markov or ARIMA models in the literature but rather the basic generation mechanism is extracted from the given observation sequence partially after the division of the variable domain into subsets with triangular fuzzy numbers. The first part of given time series is used to identify the transitional matrix elements with the given subsets. This step is referred to as the training period and helps to determine the fuzzy relationship between the temperature states at any two successive time instances. Such reletionships are expressed as conditional satatements in the form of **IF-THEN** rules between these instances. After the identification of a set of valid **IF-THEN** rules the remaining part of the series is predicted with the rule base at hand.

The application of the methodology is performed for daily maximum temperature values recorded at Kandilli observatory at the Asian coastal part of Istanbul City. The procedure yields the general statistical features of the past records including the trend and periodicity components without using any global or local periodicity-trend methodologies such as Fourier or linear trend analysis.

4. Notations

$\mathbf{A}_j\mathbf{(i)}$	=	j-th fuzzy subset of input variable at i-the fuzzy rule base,
$\mathbf{B(i)}$	=	i-th fuzzy subset of output variable,
m	=	number of implications,
R_i	=	i-th fuzzy rule base in the system,
$\mathbf{R(X, Y)}$	=	fuzzy relationship,
$\mathbf{X}_j$	=	i-th input variable for fuzzy set,
$\mathbf{Y}$	=	fuzzy output variable for the system,
$p_i(x)$	=	fuzzy basis function,
y	=	particular value of Y,
o	=	sup-star composition.

References

Box, G. E. P., Jenkins, G. M., "Time series analysis, forecasting and control", Holden-Day, San Francisco, 1970.

Jenkins, G. M. and Watt, D. G., "Spectral analysis and its applications", Holden-Day, San Francisco, 1968.

Kiszka, J. B., Kochanska, M. E., and Sliwinska, D. S., "The influence of some fuzzy implication operators on the accuracy of fuzzy model. Part I.", Fuzzy Sets and Syst., 15, 111-128, 1985a.

Kiszka, J. B., Kochanska, M. E., and Sliwinska, D. S., "The influence of some fuzzy implication operators on the accuracy of fuzzy model. Part II.", Fuzzy Sets and Syst., 15, 223-240, 1985b.

Kosko, B., "Neural networks and fuzzy systems", Prentice Hall, Englewood Cliffs, N. J., 1992.

Wang, L. X., "Stable adaptive fuzzy control of non-linear systems", IEEE Trans. on Fuzzy Systems, 1, 146-155, 1993.

Zadeh, L. A., "Fuzzy Sets", Information and Control, 8, 338-353, 1965.