A Comparison of Grey Model and Fuzzy Predictive Model for Time Series

A. I. Dounis, P. Tiropanis, D. Tseles, G. Nikolaou, and G. P. Syrcos

Abstract—The prediction of meteorological parameters at a meteorological station is an interesting and open problem. A first-order linear dynamic model GM(1,1) is the main component of the grey system theory. The grey model requires only a few previous data points in order to make a real-time forecast. In this paper, we consider the daily average ambient temperature as a time series and the grey model GM(1,1) applied to local prediction (short-term prediction) of the temperature. In the same case study we use a fuzzy predictive model for global prediction. We conclude the paper with a comparison between local and global prediction schemes.

Keywords—Fuzzy predictive model, grey model, local and global prediction, meteorological forecasting, time series.

I. INTRODUCTION

THE current weather forecasting tools, based on numerical techniques, are not always able to capture local variability in the weather. Local prediction is forecasting the future based only on a small set of the most recent data in time series. Forecasts of this kind are used to establish a curve for a most recent set of data, and then make predictions based on the established curve. In order to improve the current forecast system the ideas and algorithms of grey models are used [5]. Grey prediction can be considered as a curve fitting approach that has exceptionally good performance for real world data.

The grey system theory, first proposed by J. Deng in 1982, avoids the inherent defects of conventional methods and only requires a limited amount of data to estimate the behavior of an uncertain system or a time series. Grey means incomplete or uncertain information. The grey system has been successfully applied to industrial, social, and ecological systems, economy, geography traffic, management and environmental sciences [6,7,8,15].

In this paper we represent a fuzzy predictive model (Wang-Mendel method) for global prediction which learns an input-output mapping [Chapter 5, 16], [5], [17]. The WM method was one of the first methods to design fuzzy systems from data.

The WM method gives accurate prediction and at the same time is easy to explain to the non-expert. The method has been applied to a variety of problems [3,4,16].

The source of origin of the temperature data of the period from 1981 up to 2003 (8390 samples) was: National Observatory of Athens (NOA), Institute for Environmental Research and Sustainable Development (IERSD). The grey model and fuzzy model were implemented using MATLAB®.

The paper is organised as follows. Section 2 presents the methodology to create a predictor. In section 3 the grey modelling approach that acts as the local prediction scheme is discussed. Section 4 presents the fuzzy predictive model acting as global prediction scheme. Simulation results and the comparison of two prediction schemes are then discussed in section 5. Conclusions are made in the final section.

II. TIME SERIES PREDICTION

In general, the predicted value of a variable in a future time is based on m previous values. The m is called the lag of the prediction. If we have the values of variable y for the moments from k-m to k-1, that is, y(k-1), y(k-2), …, y(k-m), we may predict y(k), and also the next time interval values y(k+1), …, y(k+p) where p is the time step. The methodology used to train a predictor is summarized as follows:

1. Pre-process data.
2. Decide the m lag values.
3. Separate the actual data set into a training data set and a test data set.
4. Create a local or global predictor based on the architectures that follow in the next sections.
5. Use the training data set to train the predictor. The training proceeds as follows. At time k, apply y(k-1), y(k-2), …,y(k-m) to the predictor. Take the prediction output y(k+p). Calculate the output errors (criteria evaluation).
6. Evaluate the performance of the trained predictor with the test data set.

The predictor uses a set of m-tuples as inputs and a single output as the target value of the predictor. This method is often called the sliding window technique as the m-tuples slides over the full training set.

III. LOCAL PREDICTION AND GREY MODEL

Local prediction is forecasting the future based only on a set of the most recent data in a time series. In order to improve the

World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:2, No:6, 2008

current forecast system, the ideas and algorithms of grey models are used. Summary of techniques for local prediction schemes: 1) First order polynomial fitting [15], 2) GM (1,1) [6,7,8,15], 3) Exact polynomial fitting [15], 4) Fourier Grey Model (FGM) [15], 5) Linear and nonlinear exponential smoothing methods (ES) [15].

A grey system is a system that is not completely known, i.e., the knowledge of the system is partially known and partially unknown. In recent years, grey models have been successfully employed in many prediction applications. A grey modeling algorithm is described as follows.

1st step: Assume that the original raw data series $y^{(0)}$ with n samples is expressed as:

$$y^{(0)} = [y^{(0)}(1), y^{(0)}(2), ..., y^{(0)}(n)] \quad n \geq 4$$

where the superscription (0) represents the original series. We assume that the original data are positive. Negative values in data series are prohibited in grey modeling. Whenever negative values appear in the data sequence, the absolute value of the maximum negative data is added to shift all data to be positive. The task is to predict $y^{(0)}(n+p)$ where p determines the prediction sampling time, p≥1.

2nd step: Pre-processing of original raw data. The original sequence $y^{(0)}$ is transformed into a new sequence $y^{(1)}$ using the first-order Accumulated Generating Operations (AGO). AGO weakens randomness of the raw data to generate a regular sequence $y^{(1)}$.

$$y^{(1)}(k) = AGO \cdot y^{(0)} = \sum_{m=1}^{k} y^{(0)}(m), \quad k = 1,2,3,...,n \quad (1)$$

$$y^{(1)}(k) = [y^{(1)}(1), y^{(1)}(2), ..., y^{(1)}(n)]$$

3rd step: The $y^{(1)}$ sequence can be modeled by a first-order differential equation (Whitening Equation, WE) as follows:

$$\frac{dy^{(1)}}{dt} + a_g \cdot y^{(1)} = u_g \quad (2)$$

where the parameters α_g and u_g are called the development coefficient and grey input respectively. This grey model is referred to as GM(1,1), in which the first number in the brackets denotes the order of differential equation (first order) and the second indicates the number of variables (here a single variable $y^{(1)}$).

We define $z^{(1)}(k)$ as the sequence obtained by applying the MEAN operation to $y^{(1)}$

$$z^{(1)}(k) = MEAN \cdot y^{(1)} = \frac{1}{2} \cdot [y^{(1)}(k) + y^{(1)}(k-1)], \quad k = 2,3,...,n$$

Generally the mean operation may be expressed as [2]:

$$z^{(1)}(k) = ay^{(1)}(k) + (1-a)y^{(1)}(k-1)) \quad a \in [0,1] \quad (3)$$

Since the sampling time is 1 we have:

$$dy^{(1)}/dt = y^{(1)}(k) - y^{(1)}(k-1) \overset{AGO}{=} y^{(0)}(k) \quad (4)$$

By substituting equations (1,3,4) into equation (2) one has the Grey Differential Equation (GDE):

$$y^{(0)}(k) + a_g z^{(1)}(k) = u_g \quad (5)$$

In order to find the solution of the GDE (5) the parameters a_g and u_g must be solved by means of the Least Square Error Method as

$$\underbrace{\begin{bmatrix} y^{(0)}(2) \\ y^{(0)}(3) \\ y^{(0)}(4) \\ . \\ . \\ . \\ y^{(0)}(n) \end{bmatrix}}_{Y} = \underbrace{\begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ -z^{(1)}(4) & 1 \\ . & . \\ . & . \\ . & . \\ -z^{(1)}(n) & 1 \end{bmatrix}}_{B} \cdot \underbrace{\begin{bmatrix} a_g \\ u_g \end{bmatrix}}_{\Theta}, \quad \mathbf{y} = \mathbf{B} \cdot \Theta$$

$$\hat{\Theta} = (\mathbf{B}^T \cdot \mathbf{B})^{-1} \cdot \mathbf{B}^T \cdot \mathbf{y}^N$$

The solution of the Whitening Equation (2) is an exponential function and with the initial condition $y^{(1)}(0) = y^{(0)}(1)$. The solution of equation (5) can be expressed as and the solution of (2) as follows:

$$\hat{y}^{(1)}(n+p) = (y^{(0)}(1) - \frac{u_g}{a_g}) \cdot e^{-a_g \cdot (n+p-1)} + \frac{u_g}{a_g}, \quad n \geq 4$$

$n+p$ is the forecasting p step-size and $k+1 = n+p$ is the time instant of the prediction.

4th step: Take the inverse AGO (IAGO) on sequence $\hat{y}^{(1)}(k)$ we have:

$$\hat{y}^{(0)}(n+p) = \hat{y}^{(1)}(n+p) - \hat{y}^{(1)}(n+p-1)$$

$$\hat{y}^{(0)}(n+p) = (y^{(0)}(1) - \frac{u_g}{a_g}) \cdot e^{-a_g \cdot (n+p-1)} \cdot (1 - e^{a}), \quad (6)$$

World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:2, No:6, 2008

The four steps that are described above can be constructed concisely by the following scheme:

$$\hat{y}^{(0)}(n+p) = IAGO \cdot GM(1,1) \cdot AGO \cdot y^{(0)}$$

The number of data used in a grey model is small because only two parameters need to be identified. For this reason the grey approach is often used as a local predictor.

IV. GLOBAL PREDICTION AND FUZZY PREDICTIVE MODEL

Global prediction schemes employ all training data as input. Summary techniques for global prediction schemes: 1) Fuzzy predictive model [3,4,5,9,12,15,16,17]. 2) A neural fuzzy inference network SONFIN [10], 3) Case-Based Reasoning [14], 4) Adaptive Network-Fuzzy Inference System ANFIS. [10]. 5) A Genetic Fuzzy Predictor Ensemble (GFPE) for the accurate prediction of the future in time series [11]. 6) Neural Networks [1,12,13].

Fuzzy predictive model (Wang-Mendel)

The WM's method for prediction system design presents three characteristics: simplicity, a one-pass operation on the numerical input-output pairs to extract the rules and fast computational time. Suppose we are given N input–output samples:

$$\left(x_1^{(p)}, x_2^{(p)}, ..., x_M^{(p)}; y^{(p)} \right), \; y^{(p)} \in R, \; p = 1,2,...,N$$

where x_i are inputs, M is the number of the inputs and y is the output. This method consists of the following five steps.

1st step: Divide the input and output spaces into fuzzy regions

We consider that the input x_i and the output y lie in the domain intervals $\left[x_{i/min}, x_{i/max} \right]$ and $\left[y_{min}, y_{max} \right]$ respectively. We divide each interval into $2z+1$ fuzzy regions and assign each region a symmetrical triangular fuzzy set. Of course, other shapes of membership functions are possible.

2nd step: Data-generated fuzzy rules

From the training set, take the mth numerical data pair

$$\left(x_1^{(m)}, x_2^{(m)}, ..., x_M^{(m)}; y^{(m)} \right)$$

for each data pair calculate their respective membership grades in the attributed fuzzy sets. Next, choose for each variable their highest membership degree from the respective grades. Now, a rule from the m training pair is obtained:

R$^{(m)}$: **IF** x_1^m is A_1^m **AND…AND** x_M^m is A_M^m

THEN y^m is C^m \qquad (7)

where A_i^m and C^m are fuzzy sets that attributed in the condition and conclusion parts of the rule and m is the index of the rule. Especially, we define $l_i (i = 1, ..., M)$ fuzzy sets A_i^q, $q = 1, ..., l_i$ for each input x_i and l_o represent the number of membership functions in the output space. The fuzzy set A_i^m is one of the A_i^q's. Generally, in real applications we give in the fuzzy sets linguistic names like "big", "very positive", etc.

3rd step: Assign a degree to each rule

As there are usually many data pairs and therefore many rules are generated, there is high probability of conflict, that is, rules which have the same IF part and a different THEN part. To resolve this problem is to assign a truth degree (TD) to each rule and accept only the rule that has the largest truth degree. We use the following product strategy:

$$TD^{(m)} = \mu_{A_1^m}(x_1^{(m)}) \cdots \cdots \mu_{A_M^m}(x_M^{(m)}) \cdot \mu_{C^m}(y^{(m)})$$

4th step: Create a combined fuzzy rule base

The maximum number of rules that can be generated is $l_1 \cdot l_2 \cdot ... \cdot l_M$. From the 3rd step the reduction of the number rules is achieved. The generated rules determine a combined fuzzy rule base.

5th step: Determine a mapping based on the combined fuzzy rule base

Determine the overall continuous fuzzy predictive model. Using the combined rule base with K fuzzy rules in the form (7), the product inference engine, the singleton fuzzifier and the center-average defuzzifier, the following fuzzy system is obtained [17]:

$$y = f(x) = \frac{\sum_{j=1}^{K} y_c^j (\prod_{i=1}^{M} \mu_{A_i^j}(x_i))}{\sum_{j=1}^{K} (\prod_{i=1}^{M} \mu_{A_i^j}(x_i))}$$

where y_c^j is the centre of C^j. The output variable y based on the inputs $(x_1, x_2, ..., x_M)$.

V. SIMULATION RESULTS

For the local prediction of ambient time series temperature, the grey model GM(1,1) is employed. We use four data of the most recent daily average temperatures as model inputs. We know from the past research that α always equals 0.5. But with this value the error may be too large, making it unacceptable. The parameter α is critical for the grey model performance [2].

World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:2, No:6, 2008

In this paper we found the optimal value α=0.71 using Genetic Algorithms.

The performance of the forecasting models was evaluated according to criteria: Mean Square Error (MSE), Absolute Mean Error (AME) and Correlation Coefficient (ρ), as follows:

$$MSE = \frac{1}{n}\sum_{k=1}^{n}(y(k)-\hat{y}(k))^2$$

$$AME = \frac{1}{n}\sum_{k=1}^{n}\left|y(k)-\hat{y}(k))\right|$$

$$\rho = \frac{\sum_{k=1}^{n}(y(k)-\overline{y})\cdot(\hat{y}(k)-\overline{\hat{y}})}{\sqrt{\sum_{k=1}^{n}(y(k)-\overline{y})^2\cdot\sum_{k=1}^{n}(\hat{y}(k)-\overline{\hat{y}})^2}}$$

$$-1 \le \rho \le +1$$

The $y(k)$ is the actual value for time k, $\hat{y}(k)$ is the predicted value (model output) for the time k and n is the number of test data used for prediction, $\overline{y}$, $\overline{\hat{y}}$ are the mean of the actual and predicted values, respectively. The first criterion is a measure of the average squared error for all points. Correlation coefficient ρ (Pearson's formula) measures how well the predicted values correlate with the actual values. Clearly, correlation coefficient value closer to positive unity means better forecasting.

Notice that in our simulation results only the errors for the test data sets are reported because local prediction scheme do not have training errors. Temperature data are provided in Centigrade, so, apparently, they take negative or zero values. In order to employ the GM(1,1) for the local prediction, we add the absolute value of the maximum negative of the original data plus one, to each data type Let y be the original data, the transformed data will be: $y + \left|min(y)\right| + 1$

The studied Grey model scheme has the form:

$$GM(1,1): \underbrace{(y(k-1), y(k-2), y(k-6), y(k-7)}_{\text{Inputs}}; \underbrace{y(k))}_{\text{Output}}$$

The daily average temperatures (in °C) data for Athens, from 1981 to 2003 are plotted in Fig. 1. For the global prediction scheme we split the collected data into two categories. The training set consists of the temperature of the first seventeen years (1981-1997), while the test set includes the remaining six years (1998-2003). The choice of four inputs to form the WM method is case-dependent. It is an open question of how to select the optimal number of data points.

The results are tabulated in Table I. For the numerical fuzzy approach 138 rules were obtained. Table I shows that concerning the fuzzy model, the grey model has the best performance with regard to the AME criterion but the worst performance according to the other criteria. The computational time has also been recorded in the table of comparisons. The training time for WM's method is 32.366sec in Pentium M 1,7GHz, 512 MB RAM. The local grey prediction model does not have a training phase. In Fig 2 we present the results of the fuzzy and grey models for testing data for 1998 to 2003. Fig. 3 depicts the results of fuzzy and grey model for training data 1981 to 1997.

TABLE I
PREDICTION COMPARISON

Predictor	Fuzzy (WM)				Grey			
Inputs	T(k-1)	T(k-2)	T(k-6)	T(k-7)	T(k-1)	T(k-2)	T(k-6)	T(k-7)
Mfs	9	7	7	5	-			
Output	T(k)				T(k)			
Mfs	15							
AND Method	Product							
Implication	Product				-			
Aggregation	Sum							
Defuzzifier	Centroid							
Rules	138							
Results	Train		Test		Local Prediction			
Computation Time	32.366 sec		2.063 sec		0.471 sec			
AME	1.385728		1.447687		1.407747			
MSE	3.066376		3.375785		3.524036			
ρ	0.971563		0.972925		0.970235			
Maximum Error (absolute)	8.736815		8.198800		9.056811			

World Academy of Science, Engineering and Technology
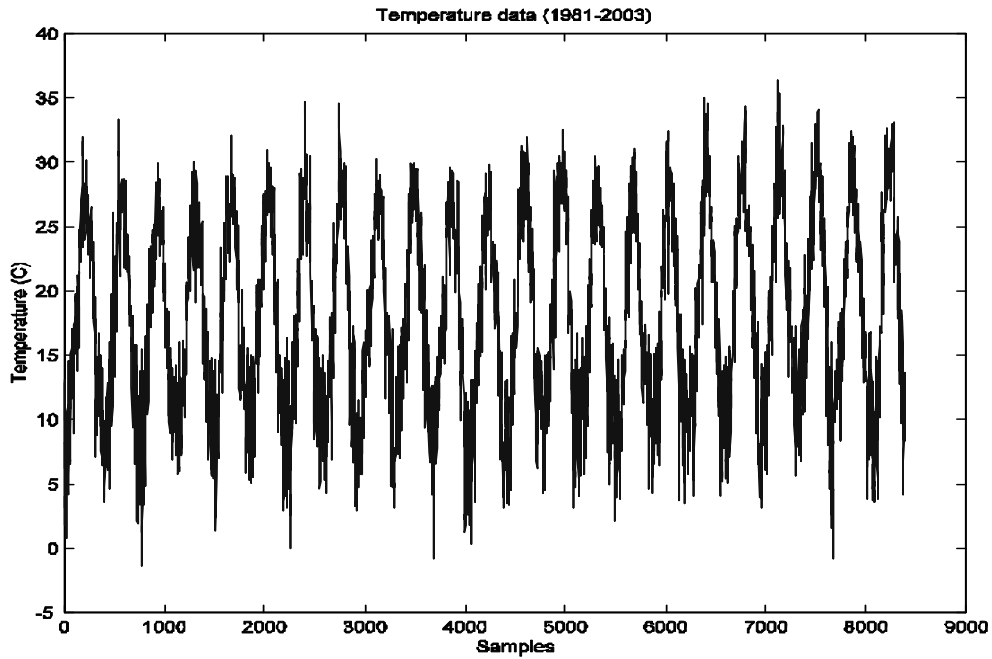International Journal of Computer and Information Engineering
Vol:2, No:6, 2008

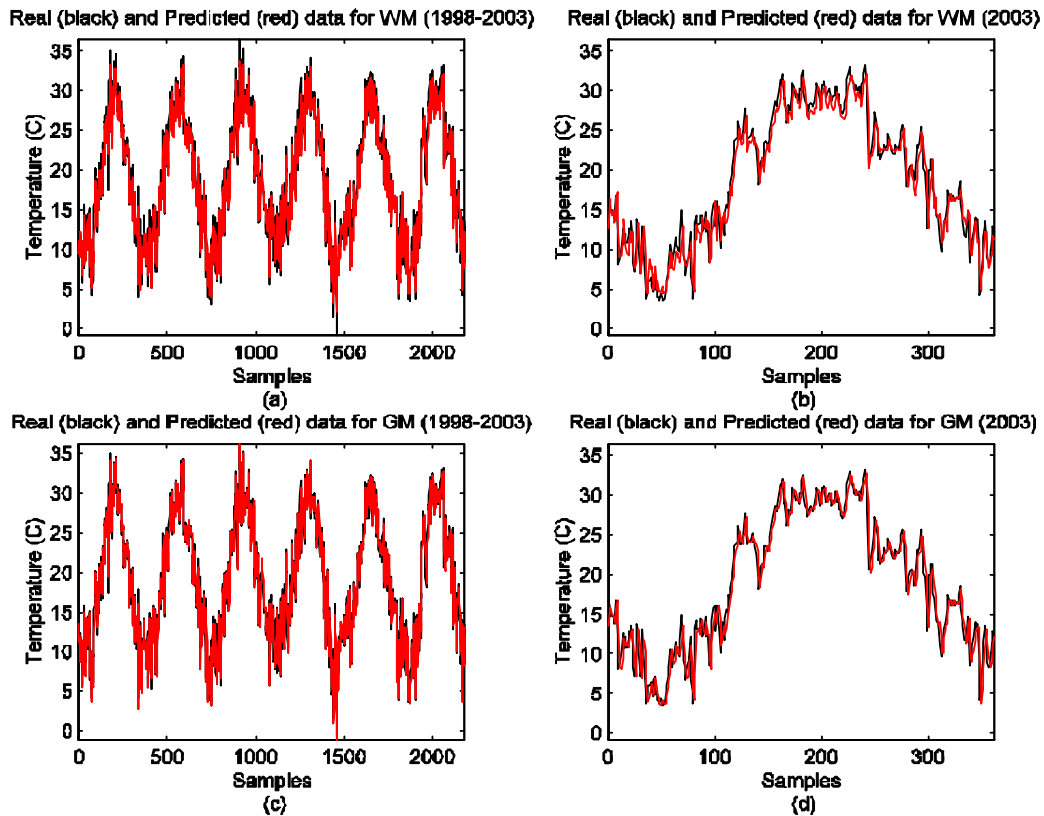Fig. 1 The daily average temperatures (in °C) for Athens, from 1981 to 2003

Fig. 2 The test results of daily temperatures between 1998 to 2003 for the WM method (a) and GM(1,1|0.71) model (c). The other two diagrams (b,d) represent the test results for the two models only for 2003 respectively

World Academy of Science, Engineering and Technology
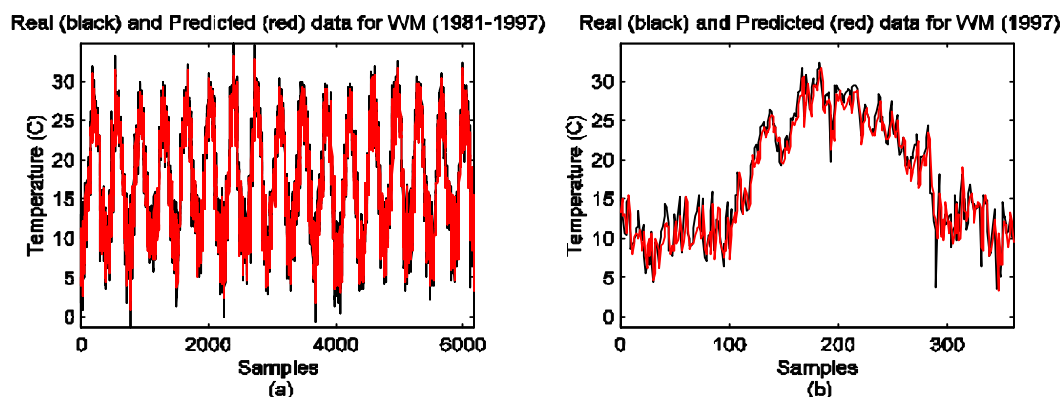International Journal of Computer and Information Engineering
Vol:2, No:6, 2008

Fig. 3
(a) The training results of daily temperatures since 1981 to 1997 for fuzzy predictive model (b)The training results only for the 1997

VI. CONCLUSION

In this paper, we have proposed a first-order linear dynamic model GM(1,1) for local prediction and a fuzzy predictive model for global prediction. From our simulation results it can be seen that the global prediction scheme has relatively better performance than the local prediction model. The results of comparison show that the knowledge from the global prediction can be incorporated in the local prediction scheme aiming at the creation of an integrated model for temperature forecasting problems.

REFERENCES

[1] A. F. Atiya, S. M. El-Shoura, S. I. Shaheen and M. S. El-Sherif, A comparison between neural-network forecasting techniques-case study: river flow forecasting, *IEEE Tran. On Neural Networks*, Vol. 10, No. 2, March 1999, pp. 402-409.

[2] T-C Chang, K-L Wen, H-T Chang, M-L You, Inverse approach to find an optimum α for grey prediction model, *1999 IEEE International Conference on SMC*, Oct. 1999, Tokyo Japan, pp. III (309- 313), 12-15.

[3] A. I. Dounis, D. L. Tseles, D. Belis, M. Daratsianakis, Neuro-Fuzzy Network for Ambient Temperature Prediction, *Neties' 97 European Conference of Networking Entities*, Ancona, 1-3 Oct. 1997.

[4] A. I. Dounis, B. Brachos, B. Stathouliss, D. L. Tseles, Neuro-Fuzzy Logic System for Mean Daily Solar Radiation, *2nd Conference Automation and Technology*, Thessalonica 2-3 October, 1998, pp. 194-199.

[5] A. I. Dounis, G. Nikolaou, D. Piromalis and D. L. Tseles, Model free predictors for meteorological parameters forecasting: a review, *1st International Scientific Conference on Information Technology and Quality*, Athens 5-6 June, 2004, A.2.3.

[6] Y.-P. Huang and C.-C. Huang, The integration and application of fuzzy and grey modeling methods, *Fuzzy Sets and Systems* 78, 1996, pp. 107-119.

[7] Y.-P. Huang and C.-H. Huang, Real-valued genetic algorithms for fuzzy grey prediction system, *Fuzzy Sets and Systems* 87, 1997, pp. 265-276.

[8] Y.-P. Huang and Tai-Min Yu, The hybrid grey-based models for temperature prediction, *IEEE SMC-B*, Vol. 27, No. 2, April 1997, pp. 284-292.

[9] J.-S. R. Jang, C.-T. Sun, E. Mizutani, *Neuro-Fuzzy and Soft Computing*, Prentice Hall, 1996.

[10] C.-F. Juang and C.-T. Lin, An On-Line Self-Constructing Neural Fuzzy Inference Network and Its Application, *IEEE on FS* Vol. 6, Feb. 1998, pp. 12-32.

[11] D. Kim and C. Kim, Forecasting Time Series with Genetic Fuzzy Predictor Ensemble, *IEEE Tr. on Fuzzy Systems*, Vol. 5, No. 4, Nov. 1997, pp. 523-535.

[12] A. Mountis, G. Levermore, Weather prediction for feedforward control working on the summer data, *5th Conference on Technology and Automation*, Thessalonici, Greece, 15-16 October, 2005.

[13] B. Reinaldo, Silveira, Shigetoshi and Sugahara, NN for local meteorological forecasting, *3rd Conference on Artificial Intelligence Applications to the Environmental Science*, AMS, Feb. 2003 D.

[14] Riordan, B. K. Hansen, A fuzzy case-based system for weather prediction, *Engineering Intelligent Systems*, Vol. 10, no. 3, 2000, pp. 139-146.

[15] S.-F. Su, C.-B. Lin, Y.-T. Hsu, A high precision global prediction approach based on local prediction approaches, *IEEE SMC-C*, Vol.23, No.4, Nov. 2002, pp. 416-425.

[16] L. X. Wang, *Adaptive Fuzzy Systems and Control Design and Stability Analysis*, PTR Prentice-Hall, 1994.

[17] L. X. Wang, The WM Method Completed: A Flexible Fuzzy System Approach to Data Mining, *IEEE on FS*, Vol. 11, No. 6, Dec. 2003, pp. 768-782.

Anastasios I. Dounis received his B.S. degree in Physics, from the University of Patras, Greece in 1983, his M.S. in Electronic Automation from the National and Kapodistrian University of Athens in 1988, and Ph.D. degree from Department of Electronic and Computer Engineering in Technical University of Crete (Chania) in 1993.

From 1986 to 1990 he was a research assistant with Institute of Informatics and Telecommunications NRCPS "Democritos", Athens. During 1993-1995, he was a part time Professor of the Department of Electronic Engineering at Hellenic Air Force Academy. He is Associate Professor under contract at the Department of Automation at the Technological Educational Institute of Piraeus since 1994.

He is member of IEEE and his current research interests include neural networks, fuzzy control and modelling, genetic algorithms, intelligent control, time series prediction and machine learning.

Dimitris Tseles received his diploma in Physics and M.S.'s in Electronic Automation and in Electronics and Radiocommunications from the National and Kapodistrian University of Athens and his Ph.D. from the Aristotle University of Thessaloniki in 1983.

He is Professor at the Department of Automation at the Technological Educational Institute of Piraeus.

His current interests include CAD/CAM/CAE, time series prediction, information technology.

George P. Syrcos received his M.Sc. in Systems Engineering in 1971 from Widener University and worked for the Boeing Company for 9 years as an Aircraft Simulations Engineer. After receiving his PhD in 1981 from Rutgers University he was appointed Assistant Professor at Rutgers University and at Wilkes University before returning to Greece, where he is presently a Professor at the Technological Educational Institute of Piraeus. He is a member of Eta Kappa Nu, Tau Beta Pi and the Technical Chamber of Greece.

Grigoris Nikolaou received his M.Sc. in Control Systems from the Department of Automatic Control of the University of Sheffield in 1998 and his PhD from the same Department in 2002. He is a research associate at the Department of Automation at the Technological Educational Institute of Piraeus. His research interests include Neural Networks and Fuzzy Systems, system identification and automatic control.