Weather Predictor for Feedforward Control working on the summer data

Alexandros Mountis and Geoff Levermore
Built Environment Research Group, School of Mechanical, Aerospace and Civil Engineering,
The University of Manchester,
a.mountis@postgrad.manchester.ac.uk – geoff.levermore@manchester.ac.uk

Abstract – **For the control of heating and cooling of buildings a weather predictor with and without limited knowledge is to be developed. This project will emphasize on the summer data and thus on the control of cooling of buildings. Traditional weather forecast by simulating the atmosphere using systems of mathematical equations which is known as Numerical Weather Prediction (NWP) is not good enough to create an optimal controller for the heating and cooling of buildings. That is why a predictor adjusted to the needs for such an engineering use, is to be created. Weather in general is a completely non linear system but studying the historical data gathered from a specific place for a huge period of time, some patterns can be identified. The problem becomes –to some extent– pattern recognition and regression task, hoping that these patterns revealed in the past information will keep repeating again in the future. Weather data is available and artificial neural networks and fuzzy logic systems are selected to be used amongst the several methods for creating models able to predict.**

Index Terms
ANNs, Fuzzy Systems, Weather Prediction

I. Introduction

A. *Purpose*

Weather prediction is important for many different reasons. While the most obvious seems to be temperature forecasting for the convenience of our lives, weather prediction is also significant to the engineering science. Temperature forecasting plays an extremely important role in short and middle-term electric load forecasting [3,14]. Temperature and solar irradiation affect the way a building is heating up or cooling down. Generally, the accuracy of the forecasting depends on parameters which one chose for it. For instance, considering electric load forecasting in this manner, it seems that the result of the temperature prediction, one of the parameters for the forecasting has great influences on electric demand [5]. Furthermore almost 50% of energy in the UK is used in buildings [3]. So, huge amounts of energy can be saved by being able to predict the weather. For example if it is known that is going to be hot the next day, instead of using air-conditions after the building is occupied, natural ventilation can take place during the night.

B. *Types of weather prediction*

Weather prediction can have many different forms, depending on the required use of the prediction. For example in airports it is far more important to know about the visibility a few hours ahead rather than the temperature. Sailors care more about winds, etc. A variation also stands about how far in time would the prediction be [2]. There are a lot of occasions that we don't actually care about tomorrow's weather or what the temperature would be in a few hours ahead, but we want to know the forecast about a week's time or even more. Another issue for weather forecasts is the area that such a prediction would cover. Thus, it is easily understood that the weather predictors that have been studied so far, are created depending on the specific requirements each time. A weather predictor that would work efficiently under all the circumstances and would be able to predict all kinds of natural phenomena and give accurate predictions both for large areas and small-scale environments is practically not realizable.

C. *Aim*

The aim of this project is to develop a weather prediction system which will produce an accurate forecast. It will predict maximum and minimum daily temperatures and maximum daily global solar irradiation. The ideal system would consist of a computer and meteorological instruments attached to it. These instruments would feed real time meteorological data to the computer. The data would be analyzed and learned by the computer. In this way, the workstation could be taught about the local microclimate and thus give small scale forecasts. However, in this case, instead of using meteorological instruments to receive real time information, historical data shall be used. The study will look into the differences of the fuzzy logic with the artificial neural network approach and will suggest a predictor using a combination of these two techniques. However, the solution should not be considered as a neuro-fuzzy approach since the architecture of neuro-fuzzy systems is different from the model that combined these two methods [6]. The predictions that would be most useful for controlling a building's ventilation systems are daily temperatures and solar irradiation [3].

D. *Terminology*

Due to the peculiarity of the data set, some expressions have slightly different meanings than one would expect. For example, what is called year in this paper is different from the normal calendar year. A "year" here consists of six months. Summer months are the months from April to September. When an element is prorogued as "today's" means that it is the observation at time *t*, whereas "tomorrow's" is the prediction at time *t+1*, and "yesterday's" is the observation at time *t−1*. A time unit *t* for this project is a whole day. The temperature has units of degrees Celsius (^oC). Irradiance is a measure of the rate of energy received per unit area, and has units of Watts per square meter (W/m^2), where 1 Watt (W) is equal to 1 Joule (J) per second (s).

II. Fundamental Concepts

The fundamental concepts behind this project are fuzzy logic systems and artificial neural networks. Fuzzy systems and neural networks have attracted the growing interest of researches in various scientific and engineering areas. The area which this study is looking into is prediction. Although the fundamental inspiration for these two fields is quite different, the fact that are suitable of solving many of the same problems point out some similarities. They share the common ability to improve the intelligence of systems working in an uncertain, imprecise and noisy environment [1]. They also try to simulate the human behavior of dealing with uncertainty and decision making. In this chapter the basic ideas are briefly explained behind these two fields.

A. *Fuzzy Logic*

Fuzzy logic is based on the way the brain deals with inexact information. In other words it is extending the classical two-valued Boolean logic by providing a certain degree to uncertainties. Mathematically, it is based on the idea of fuzzy sets that can be used to model linguistic terms. In fuzzy logic, it is possible to formulate fuzzy rules that are determined by linguistic variables and apply them in order to infer a resolution [7,8]. The accuracy of a fuzzy system depends on how well its parameters are defined.

B. *Artificial Neural Networks*

Artificial neural networks are modelled after the physical architecture of the brain. They consist of a number of independent simple processors (neurons). These neurons are highly interconnected via weighted routes. The main property of an ANN is the ability to get trained by various algorithms that adjust these connection weights. By doing so, ANNs can learn, recall and generalize from training patterns and data. A great disadvantage of ANNs is that they are referred as "black boxes" because it is usually not possible to extract explicit knowledge from them [6,9]. Two or more of the neurons can be combined in a layer, and a particular network could contain one or more such layers creating multilayered networks.

III. Data Analysis

A. *Processing the data*

In today's industrial processes, there is no shortage of "information". No matter how small or how straightforward a process may be, measuring instruments thrive [10]. As portrayed, the most important role in regression is the meaningful data analysis. The given information was statistical data (purely numerical data by observations) and there was no accumulated judgment and expertise of key personnel. This historical data consisted of hourly collections of Manchester International Airport's weather measurements for the years 1982 to 1995. Since this work examines only the summer prediction, the measurements of the months October, November, December, January, February and March were completely ignored. This elimination still left 4392 hourly readings per year (183 days x 24h per day = 4392 readings). Each reading consisted of 19 elements. Not all of these elements were significant and so some were ignored. On the other hand, some new meaningful ones were derived. These were the hourly change in dry-bulb temperature, the hourly change in atmospheric pressure and the hourly change in wind speed. In this case the "hourly change" can be interpreted as the first derivative of each measurement. Some of the data had to be excluded of any analysis and not to be taken into consideration at all when searching for a suitable model, in order to have a good set to test whether each model is actually working [8,11]. Thus, 2 years, which are about 15% of the whole data set, were excluded of any analysis. A block diagram of the route to distinguishing the data can be seen in Fig 1.Since the prediction must be next day's maximum and minimum temperature and maximum solar irradiation, hourly data is practically useless. Therefore, the relevant statistics had to be calculated from the hourly data set and are as follows:

1. daily maximum dry-bulb temperature
2. daily minimum dry-bulb temperature
3. daily mean dry-bulb temperature
4. daily maximum wind speed
5. daily minimum wind speed
6. daily mean wind speed
7. daily mean cloud coverage
8. daily mean global solar irradiation
9. daily daytime mean global solar irradiation
10. daily maximum global solar irradiation
11. daily sum of global solar irradiation

12. daily maximum atmospheric pressure
13. daily minimum atmospheric pressure
14. daily standard deviation of the hourly differences of the atmospheric pressure
15. daily mean of the hourly differences of the atmospheric pressure
16. daily standard deviation of the hourly differences of the dry-bulb temperature
17. daily mean of the hourly differences of the dry-bulb temperature
18. daily standard deviation of the hourly differences of the wind speed
19. daily mean of the hourly differences of the wind speed
20. daily average wind direction

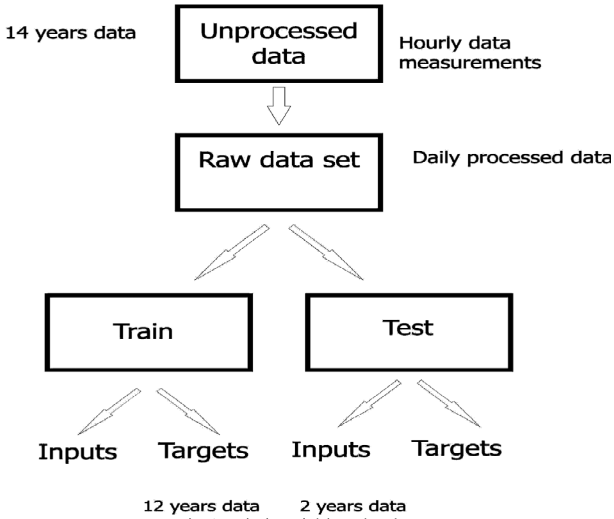

Fig 1: Distinguishing the data

B. Analyzing the data

The single most important thing to do when first exploring the data is to visualize the data through graphs. The basic features of the data including patterns and unusual observations are most easily seen through graphs. Sometimes graphs also suggest possible explanations for some of the variation in the data [11,12].

1) Time Plots: For the specific data the most obvious graphic form would be a time plot in which the data of some elements are plotted over time. In Fig. 2 one can observe that the data of the elements that will be used as outputs can be categorized as seasonal [12]. Every 182 days, the shape of the plot is roughly repeated. Especially for the temperature graphs, it starts from low temperatures for every April and gradually rises to its maximum values around July and then starts dropping again until September. In fact, daily temperatures follow approximately a sine wave, at least for the summer period, since the winter data is not being considered. Especially for the global solar irradiation, the seasonality is distinguished but not as clear as the temperature graphs. This is because global solar irradiation is affected by far more factors than temperature is at least during summer and the graph appears with "noise". Time plots though are not the only graphs that will lead to conclusions about the data set.

2) Scatter Plots: Such a plot will provide information about the relationship between the two variables presented on the graph [12]. In Fig. 3 one can distinguish a representative example. Fig. 3 is a scatter plot of the daily mean global solar irradiation over the daily maximum dry-bulb temperature. It is clearly discerned that as the daily maximum dry-bulb temperature increases so does the daily mean global solar irradiation. Hence, those two elements are highly correlated and in a way can be described as proportional.

3) Autocorrelation Plots: Since, the data can be characterized as time series as it consists of a sequence of observations, another important statistic that has to be analyzed is the autocorrelation in the elements that going to be predicted. Therefore, it is important to see for example how consecutive observations are related. If Yt is defined as the observation at time *t* then the observation Yt-1 is described as "lagged" by one period. Similarly, it is possible to describe observations lagged by two periods, three periods and so on. Together, the autocorrelations at lags 1, 2 and so on, make up the autocorrelation function or ACF [12]. The ACF graph helps us identify if previous values of the series contain much information about the next value or whether there is little relationship between one observation and the next. Fig 4. displays the ACF of daily maximum dry-bulb temperature for 20 lags [13]. We are mainly interested for the first lag correlation, which is remarkably high. Throughout this data analysis several statistics like minimum, maximum, mean and standard deviations for all the elements were gathered. Gathering these numerical summaries was essential for constructing the fuzzy sets that could describe as best as possible the given data.

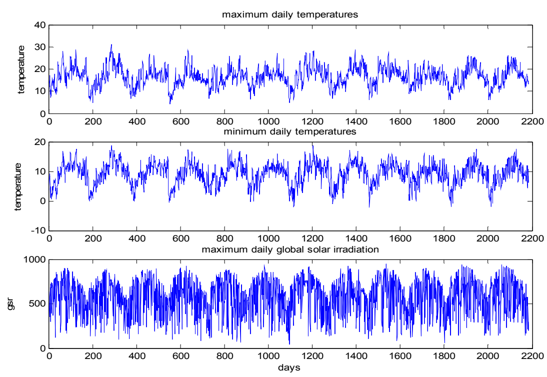

Fig. 2: Time plots of Manchester's daily maximum and minimum temperatures and maximum solar irradiation.

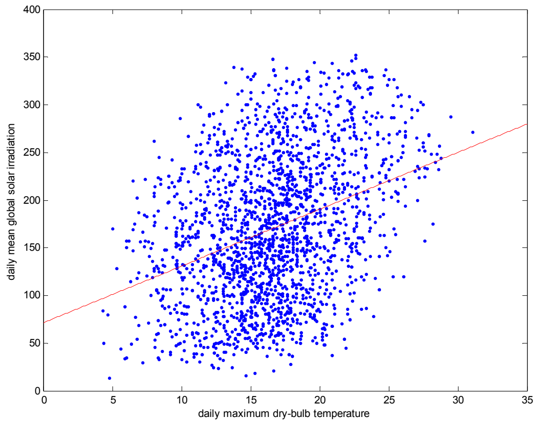

Fig. 3: Scatter plot of daily mean global solar irradiation over daily maximum dry-bulb temperature (correlation).

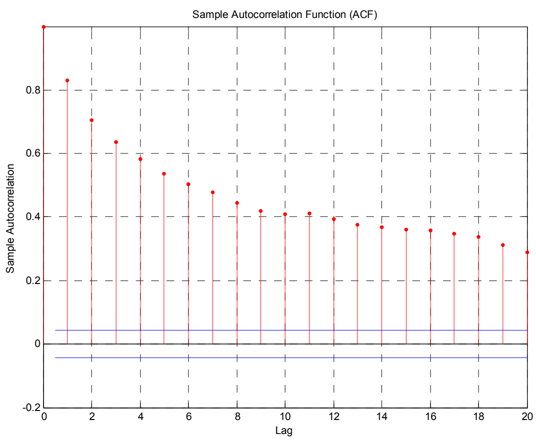

Fig. 4: ACF graph of daily maximum dry-bulb temperature.

IV. Approaches

Several models were created and several techniques were tried to achieve useful results. From the data analysis June proved to be a highly unstable month in terms of temperature and in general it can be characterized as a month with a variety of phenomena. So a model built on June data would probably be very similar to a model built using the whole data set. Using such a smaller data set would provide quickly the first readings on how the models should be approached. The target though is to build a predictor capable of providing forecast for a wider range of time. These approaches can be considered more as exploratory of how each system reacts to the data rather than a proper predictor. Having in mind the experience of the June models, the first model trying to provide a general prediction is using only fuzzy logic. The results from the fuzzy predictor were quite unsatisfactory, while the results produced by the ANN using only Junes were promising enough that once the whole data set is presented to a suitable network a better prediction would be possible. With the proper adjustments the ANN predictor provided satisfactory results. But since the weather can be portrayed as a vague concept a combination of the effectiveness of ANNs and the vagueness of fuzzy systems might prove more robust. The basic idea is that the ANN will try to fine-tune the predicted values of the fuzzy system. Amongst the 20 different potential elements of the data set, only 9 were used for this fuzzy model as inputs (see Fig 5) and are: daily maximum dry-bulb temperature, daily minimum dry-bulb temperature, daily mean dry-bulb temperature, daily maximum wind speed, daily mean wind speed, daily mean cloud coverage, daily maximum global solar irradiation, daily maximum atmospheric pressure and daily minimum atmospheric pressure. The outputs are: next day's daily maximum dry- bulb temperature, next day's daily minimum dry-bulb temperature and next day's daily maximum global solar irradiation. The fuzzy sets for the inputs and outputs would describe 5 linguistic variables. These variables are "very low", "low", "normal", "high" and "very high". The variables of "low", "normal" and "high" are described by Gaussian functions, while "very low" is a Z-shaped and "very high" is an S-shaped sigmoidal function. The membership function plots of all the elements (inputs and outputs) were created by an automated procedure, where statistics – maximum, minimum and mean value – of each element were calculated and used as parameters to the functions with an appropriate adjustment each time in order to create the representative fuzzy sets. The next step was the development of the fuzzy rules. Since there is no expert knowledge, the statistic approach for generating the rules was used. The train data set is characterized as the observed data and would be responsible for the fuzzy rules. The mean values of the elements act as the basis for categorizing the input values into linguist variables. A simple algorithm was used. The differences of the actual from the mean values determined the linguistic variable of the elements. By doing so, a total of 2184 rules were observed, of which 1721 were different. It is easily understood that a fuzzy system can not have 1721 rules. This is the reason for developing an algorithm that would weight the rules. Then the rules can be sorted out by their weight and the strongest ones can be selected to participate in the fuzzy system. The algorithm is based on the correlation matrix. The algorithm looks to all the rules and how correlated are the 3 outputs to the 9 inputs respectively. If they are highly positive correlated then the elements that are considered, must have the same fuzzy linguistic variable. If this is happening then this rule is being rewarded with a bonus (the value of which is proportional to the correlation value of the two elements). If there is a small dissimilarity in the linguistic variables then the bonus is smaller and then if the difference is high in their linguistic variables then the bonus is 0. On the contrary if the elements are highly negative correlated the procedure is reversed and bonus is rewarded to a rule if the elements have big dissimilarity in their linguistic variables. Hence, the weight of the bonus each rule is rewarded depends on both the correlation coefficients between the elements and the degree of dissimilarity of the linguistic variables. 181 rules are used and then the fuzzy model produces its output. Then, the predicted values are subtracted from the input values and the outcome is fed amongst other inputs to the ANN (see Fig 6-Remark 1). These are the elements of the whole data set that did not take part in the fuzzy system. By doing so, 19 of the 20 daily elements are actually taken into consideration, providing with as much input detail to the final model as possible. We also feed the differences with yesterday's values (see Fig 6-Remark 2). The output of the ANN predicts differences with today's inputs. This idea is based on the fact that tomorrow's readings are actually same as today's with a small error and the network this time is trying to estimate this error rather than the actual value. The architecture is 3 layers, with tan-sigmoidal transfer functions for the input and hidden layer and linear for the output layer. In this model, the input neurons are 19 the output neurons 3. The size of a hidden layer is a fundamental question often raided in the application of multilayer feedforward networks to real-world problems. The exact analysis of this issue is rather difficult because of the complexity of the network mapping and the nondeterministic nature of many successfully completed training procedures. Hence the size of a hidden layer is usually determined experimentally and by empirical guidelines [9]. Such a guideline is as follows. For a network of a reasonable size, the size of hidden nodes needs to be only a fraction of the input layer. A simple rule of thump for this guideline is the following [1]:

$$h = (m+n)\backslash 2 + \{...0...1...2...\} \quad (1)$$

where the symbol "\" denotes integer division, h the number of neurons in hidden layer, m the number of neurons in input layer and n the number of neurons in output layer. The $+\{...0...1...2...\}$ part of the equation means that if the network fails to converge to a solution with the result of the integer division, it may be that more hidden neurons are required. Consequently, by applying equation (1) and testing, the hidden neurons are 12.

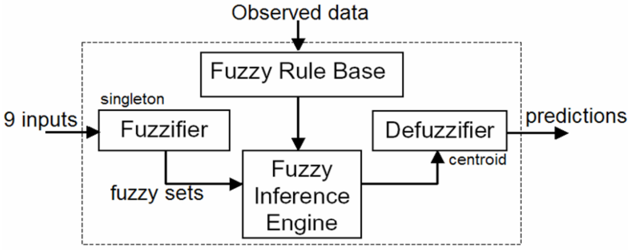

Fig 5: The fuzzy system.

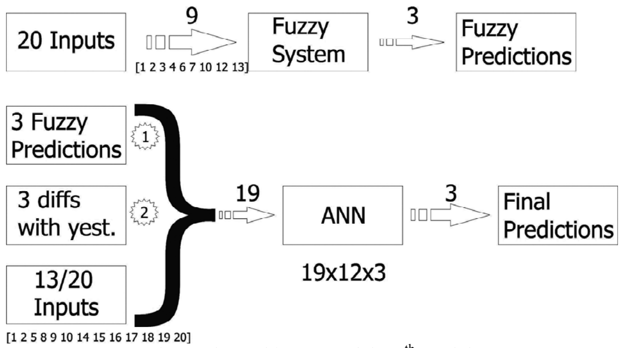

Fig 6: The architecture of the 5th model.

V. RESULTS & CONCLUSIONS

A. Measuring the accuracy

If O_t is the actual observation for time period *t* and P_t is the forecast for the same period, then the error is defined as:

Error: $$e_t = O_t - P_t \quad (2)$$

Since there are observations and predictions for *n* time periods, then there will be *n* error terms, and the following standard statistical measures can be defined:

Mean error: $$\mathrm{ME} = \frac{1}{n}\sum_{t=1}^{n} e_t \quad (3)$$

Mean absolute error: $$\mathrm{MAE} = \frac{1}{n}\sum_{t=1}^{n} |e_t| \quad (4)$$

Mean squared error: $$\mathrm{MSE} = \frac{1}{n}\sum_{t=1}^{n} e_t^2 \quad (5)$$

Equation (2) can be used to compute the error for each period. These can then be averaged as in equation (3) to give the mean error. However, the ME is likely to be small since positive and negative errors tend to offset one another. The ME was calculated but since it does not give much indication as to the size of typical errors it was finally ignored. Therefore, the MAE is defined by first making each error positive by taking its absolute value and then averaging the results. A similar idea is behind the definition of MSE. Here the errors are made positive by squaring each one, and then the squared errors are averaged. The MAE has the advantage of being more interpretable and easier to explain to non-specialists. The MSE has the advantage of being easier to handle mathematically [15]. Each of these statistics deals with measures of accuracy whose size depends on the scale of the data. That is why percentage error measures can be defined. First, the percentage error is defined as:

Percentage error: $$PE_t = \left(\frac{O_t - P_t}{O_t}\right) \cdot 100 \quad (6)$$

Then the following two relative measures are identified:

Mean percentage error: $$\text{MPE} = \frac{1}{n}\sum_{t=1}^{n} PE_t \quad (7)$$

Mean abs percentage error: $$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n} |PE_t| \quad (8)$$

Equation (6) can be used to compute the percentage error for any time period. These can then be averaged as in equation (7) to give the mean percentage error. However, as with the ME, the MPE is likely to be small since positive and negative PEs tend to offset one another. Hence the MAPE is defined using absolute values of PE in equation (8). On the other hand, the MAPE is only meaningful if the scale has a meaningful origin. However, MAPE was not used for assessing the accuracy of temperature forecasting because 1^{o}C error forecast produces significantly higher MAPE in low temperatures (predicting the minimum dry-bulb temperature) than in high temperatures (predicting the maximum dry-bulb temperature) [12].

B. Discussion & conclusions

The selected tool to measure and compare the accuracy finally was the MAE and the MSE. The MAE was selected because it would give a meaningful indication for the error and the MSE because is the most common performance function and very widely used especially in ANNs. First of all, a concentrated table (I) is presented with all the results The most accurate model for all 3 predictions is definitely the last one which uses a combination of ANNs and Fuzzy Logic Systems.

TABLE I
RESULTS

			next day's daily maximum dry-bulb temperature	next day's daily minimum dry-bulb temperature	next day's daily maximum global solar irradiation
MAE	fuzzy	1st model	**3.0306**	**3.5933**	**164.25**
MSE			15.962	19.584	39118
MAE	ANN	2nd model	**2.0717**	**1.3144**	**152.82**
MSE			6.3051	2.8688	34440
MAE	fuzzy	3rd model	**4.0206**	**5.2999**	**161.22**
MSE			24.659	44.62	42645
MAE	ANN	4th model	**1.8517**	**1.3844**	**142.94**
MSE			5.6256	2.9774	33742
MAE	both	5th model	**1.7096**	**1.1914**	**136.83**
MSE			4.7517	2.2632	31711

Although the accuracy differences with the 4th and 2nd model in which they both use ANNs as a prediction method are not that significant, one can say that a considerable improvement took place. First of all, one can observe that the models which use fuzzy systems predict better the next day's daily maximum dry-bulb temperature rather than next day's daily minimum dry-bulb temperature. None the less, the difference in accuracy with the ANN counterpart models is quite significant. However, the difference in accuracy of the next day's daily maximum global solar irradiation is not equivalent to the other differences. Hence, one can say that the models built with fuzzy logic had a successful prediction only with the next day's daily maximum global solar irradiation. If the two models built with the fuzzy logic are compared alone, then it is noteworthy to mention that the model created using the whole data set provides worse prediction in two out of the three outputs. As derived from the data analysis, the time plot of maximum global solar irradiation demonstrated less "clear" periodicity than the time plots of the temperatures. Furthermore, although the rules for the system with the whole data set were selected after being weighted, they failed to provide the model with as many qualitative rules as the ones produced by only looking at the June data. So, as a conclusion one can say that fuzzy logic is more suitable for predictions when the data exhibits high noise or little to none pattern and that the presence of qualitative and expert knowledge for producing the rules is essential. On the other hand, ANN models proved very good accuracy on the temperature predictions and a satisfactory on global solar irradiation prediction. As mentioned before it was better than the one predicted with the fuzzy models but not as better as the temperature predictions. Another general observation looking only at the ANN models is that they predict better the next day's daily minimum dry-bulb temperature rather than the next day's daily maximum dry-bulb temperature. Keeping in mind the data analysis, this was a highly expected situation, since the daily minimum dry-bulb temperature proved to follow a pattern more than the daily maximum dry-bulb temperature. The important thing was that when the model was introduced to more data the accuracy of the predictions was considerably improved. So, as a conclusion one can say that ANNs are excellent for predictions when there is a pattern in the data and that the more data introduced to the network the better the results turn out to be. Furthermore, expert knowledge has little to none significance. Consequently, since the data which was introduced to the last two models (the only ANN model and the one with the combination of the fuzzy and ANN model) was of the same quantity and information one should expect the same accuracy. Up to a point this is the case. The results of the model built with the combination of fuzzy system and ANN showed little improvement. One can say that the ANN with the specific architecture (which is the most widely used in regression problems) can not improve anymore. However results did show a slight improvement and this is mainly because the architecture was adjusted to be used after a fuzzy system and 3 new elements were brought into the network. The improvement though happened due to the nature of these 3 new elements which helped the network to produce more accurate predictions. Fig. 7 presents graphically the deviation in the predicted values over the observed values for some selected days of April 1990.

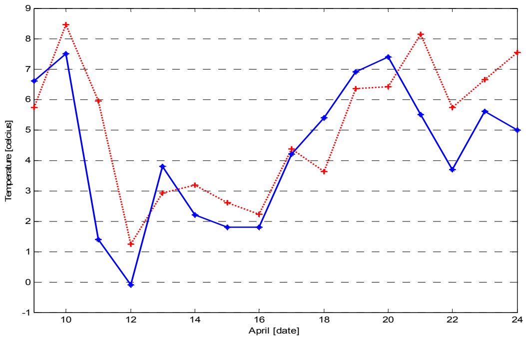

Fig. 7: Minimum daily temperature of selected days of the April (Red crosses, observed - Blue stars, forecasted).

REFERENCES

[1] Masayuki Nagano and Bahman Kermanshahi, "Temperature forecasting using artificial neural networks," The ICEE Proceedings in Kitakyushu, July 2000

[2] Peter Bennett, "A neural net-based weather prediction system," Project Report, Knowledge-Based Systems MSc, School of Cognitive and Computing Sciences, University of Sussex, August 1996

[3] G. J. Levermore, "Building energy management systems: applications to low-energy HVAC and natural ventilation control," 2nd ed., London: E & FN Spon, 2000

[4] Abdul Manan Ahmad, Chia Su Chuan and Fatimah Mohamad, "Weather Prediction Using Artificial Neural Network," The International Conference on Circuits/Systems Computers and Communications (ITC-CSCC 2002) Proceedings, 2002, pp 262-264

[5] Krzysztof Siwek and Stanislaw Osowski, "Regularization of neural networks for improved load forecasting in power system," IEEE International Conference on Electronics, Circuits and Systems ICECS 2001, Malta 2001, Vol. II, pp. 1255-1258

[6] Detlef Nauck, Frank Klawonn and Rudolf Kruse, "Foundations of neuro-fuzzy systems," New York; Chichester: Wiley, 1997

[7] Robert E. King, "Υπολογιστική νοημοσύνη στον έλεγχο συστημάτων," 1η έκδ., Αθήνα: Π. Τραυλός – Ε. Κωσταράκη, 1998

[8] Janusz Kacprzyk and Mario Redrizzi, "Fuzzy regression analysis," *Series studies in fuzziness*, vol. 1, Warsaw: Omnitech Press, 1992

[9] Chin-Teng Lin and C.S. George Lee, "Neural fuzzy systems: a neuro-fuzzy synergism to intelligent systems," Upper Saddle River; NJ: Prentice-Hall, 1996

[10] Norman R. Draper and Harry Smith, "Applied regression analysis," 3rd ed., New York; Chichester: Wiley, 1998

[11] Vladimir Cherkassky and Filip M. Mulier, "Learning from data: concepts, theory and methods," New York; Chichester: Wiley, 1998

[12] Spyros Makridakis, Steven C. Wheelwright and Rob J. Hyndman, "Forecasting methods and applications," 3rd ed., New York; Chichester: Wiley, 1998

[13] Mathworks, Matlab Help [2004], http://www.mathworks.com/

[14] American Society of Heating, Refrigerating and Air-conditioning Engineers website [2004], http://www.ashrae.org/

[15] Martin Brown, "Lectures notes, CSC: intelligent systems," Manchester, 2004