Elman NN and Time Series in Forecasting Models for Decision Making

Conference Paper · August 2006

DOI: 10.1109/WAC.2006.376047 · Source: IEEE Xplore

4 authors:

Cvetko Andreeski
University "St. Kliment Ohridski" - Bitola
26 PUBLICATIONS 44 CITATIONS
SEE PROFILE

Pandian Vasant
Universiti Teknologi PETRONAS
218 PUBLICATIONS 1,518 CITATIONS
SEE PROFILE

Mile Stankovski
Ss. Cyril and Methodius University
79 PUBLICATIONS 123 CITATIONS
SEE PROFILE

Georgi Dimirovski
Dogus Universitesi
343 PUBLICATIONS 2,216 CITATIONS
SEE PROFILE

Correspondence Author: Cvetko J. Andreeski, cvetko.andreeski@uclo.edu.mk
Draft Paper

ELMAN NN AND TIME SERIES IN FORECASTING MODELS FOR DECISION MAKING

Cvetko J. Andreeski [1], Pandian M. Vasant[2], Mile J. Stankovski [3], and Georgi M. Dimirovski[4]*

[1] St. Clement of Ohrid University - Bitola, Faculty of Tourism and Hospitality - Ohrid
Key Marsal Tito 95, MK-6000 Ohrid, Republic of Macedonia
Phone: ++389-46-262247; Fax: ++389-46-264215
E-mails: cipuslju@mt.net.mk; cvetko.andreeski@uclo.edu.mk

[2] University Teknologi Petronas, Electrical & Eelctronic Engineering
31750 Tronoh, BSI, Perak DR, Kingdom of Malaysia.
E-mail: pandian_m@petronas.com.my

[3] SS Cyril & Methodius University, Faculty of Electrical Engineering
Karpos 2 BB, MK-1000 Skopje, Republic of Macedonia
E-mails: milestk@etf.ukim.edu.mk

[4] Dogus University, Faculty of Engineering
Acibadem, Kadikoy, TR-34722 Istanbul, Turkey
E-mail: gdimirovski@dogus.edu.tr

Abstract – This paper examines and compares analytical tools in analysis of economic statistical data, econometric modeling, and neural network, soft-computing modeling, as representation models for time series processing in forecasting, decision and control. In addition, a novel forecasting model using Elman networks is proposed. A comprehensive experiment in applying the latter modeling has been carried out, some specific applications software developed, and a number of benchmark series from the literature processed. This paper reports on comparison findings as well on the use of our application software package which encompasses routines for regression, ARIMA and NN analysis of time series. The comparison analysis is illustrated by a sample example known as difficult to model via any technique.

Keywords: Analysis of time series; decision; forecasting; neural networks; patterns.

1. INTRODUCTION

Many sets of data appear in the form of time series: monthly sequence of number of tourists who visit a given resort; hourly observations on the yield of a chemical process; a weekly sequence of the number of road accidents, and so on. Examples of time series are abound in such fields as business management, engineering, finance management, natural sciences. Time series analysis is concerned with techniques for the analysis of the dependence between adjacent observations. This requires the development of stochastic and dynamic models for time series data and the use of such models in application areas of concern. These models, in turn, give analytic results in terms of system model (e.g. transfer function) that identifies cause-

* This research is supported in part by Faculty of Tourism and Hospitality in Ohrid, and in part by Dogus University Istanbul.

effects relationship some given time series for more general study of the underlying phenomena for various purposes, including forecasting, decision and control. Such models can be used for several reasons:

- Analyzing dependence between observations or involvement of some other time series in the observed one.
- Predicting future values for the observed time series (forecasting) in oreder to facilitate decision-making.
- Designing of practical and simple control scheme by means of which potential deviations of the system output from a desired target may be compensated by adjustment of the input series values so far as possible.

These basic reasons are indeed essential in economic analysis, and that is why they should be part of our education.

Classical pattern recognition has mainly been concerned with detecting systematic patterns in an array of measurements, which do not change in time (static patterns). Typical applications involve the classification of input vectors into one of several classes (discriminant analysis) or the approximate description of dependencies between observables (regression). These applications use linear models for identification and prediction.

2. FOR TIME SERIES PROCESSING USING NEURAL-NET MODELS

It is known that different neural networks, according to the type of mechanism to deal with time series, can be used. Most neural networks have previously been defined for pattern recognition in static patterns; the temporal dimension has to be supplied additionally in an appropriate way. Some of them are worth mentioning [2]: layer delay without feedback (time windows); layer delay with feedback; unit delay without feedback; unit delay with feedback (self-recurrent loops).

In this research the class of recurrent network created with adding new elements in classic neural feed-forward networks, which can be trained during a certain sequence of time, have been studied. The important feature of kind of neural networks is that they can identify dynamic systems with no need for more than one previous value of the input and output. Therefore these are capable of identifying dynamic systems with unknown order or with unknown time delay. This network can be classified as an Elman network.

A common method for time series processing are so called (linear) state space models [3]. The assumption is that a time series can be described as a linear transformation of a time dependent state-given through a state vector $\vec{s}$:

$$\vec{x}(t) = C\vec{s}(t) + \vec{\varepsilon}(t) \tag{1}$$

where C is some transformation matrix. A linear model usually describes time dependent state vector:

$$\vec{s}(t) = A\vec{s}(t-1) + B\vec{\eta}(t) \tag{2}$$

where A and B are matrices, and $\vec{\eta}(t)$ is a noise process, just like $\vec{\varepsilon}(t)$ above. The model for the state change, in this version, is basically an ARMA[1,1] process. The basic assumption underlying this model is the so-called Markov assumption [ ]: the next sequence element can be predicted by the state a system producing the time series is in, no matter how the state was reached. In other words, all the history of the series necessary for producing a sequence element can be expressed by one state vector. Should the assumption that the states are also dependent on the past sequence vector holds, by neglecting the moving average term B $\vec{\eta}(t)$, one obtains

$$\vec{s}(t) = A\vec{s}(t-1) + D\vec{x}(t-1)\,. \tag{3}$$

Then, basically, an equation describing a type of recurrent neural network known as Elman network [2] (depicted in Figure 1) is obtained. The Elman network is in fact a multi layer perceptron (MLP) neural-net computing structure with an additional input layer, called the state layer, receiving as feedback a copy of the activations from the hidden layer at the previous time step. Should use this network is employed for forecasting modelling, the activation vector of the hidden layer is equated with $\vec{s}$, and then the only difference to Eq. (3) is the fact that in an MLP a sigmoid activation function is applied to the input of each hidden unit

$$\vec{s}(t) = \sigma(A\vec{s}(t-1) + D\vec{x}(t-1)) . \quad (4)$$

Here σ(a) refers to the application of the sigmoid (or logistic) function $1/(1+\exp(-a_i))$ to each element a_i of A. Hence the transformation is no longer linear but a non-linear one according to the logistic regressor applied to the input vectors.

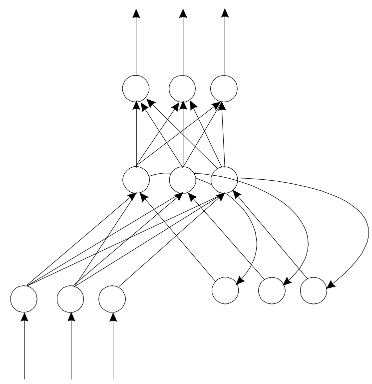

Figure 1. The structure of Elman neural networks

The Elman network can be trained with any learning algorithm that is applicable to MLPs such as backpropagation (implemented in our application software) or conjugant gradient, to name a few. This network belongs to the class of so-called simple recurrent networks. Even though it contains feedback connection, it is not viewed as a dynamical system in which activations can spread indefinitely. Instead, activations for each layer are computed only once at each time step (each presentation of one sequence vector).

3. BOX-JENKINS APPROACH AND NLARX MODELS

Many economic and financial time series observations are nonlinear; hence linear parametric time-series models may fit data poorly (see below). An alternative approach, which implements non-linear models, is via the use of artificial neural networks [1], [4]. We can point two basic characteristics that make them very attractive for time series prediction: the ability to approximate functions, and the direct relationship with classical models, such as Box-Jenkins models. Despite these advantages, there are problems to be observed in using neural networks. ANN computing models are, in general, more complex and involved than linear models, hence more difficult to design. Furthermore, they are more vulnerable to the problems of overfitting and local minimum [4].

The NARX model (Non-linear AutoRegressive model with eXogenous variables) is described by Eq. (4). This class of models can be implemented by all four types of neural-net computing structures, which can be used for modelling time series of economy nature and origin. In the sequel we focus on modelling of a time series with both models: Box-Jenkins model and Artificial Neural Network (NARX) model. The sample time series used is taken from Box-Jenkins' book "Time Series Analysis: Forecasting and Control", and represents - Series B IBM common stock closing prices: daily [3]. In order to get a stationary time series equivalent, for the modelling with ARIMA model (Figure 1) we made one differencing and one seasonal differencing. For the modeling with NARX ANN, we made one differencing and normalization in the interval [-1, 1]. The number of the dependent variables for the ARIMA model is determined by Bayesian-Schwarz criteria. Results on the grounds of this criterion are given in Table 1. From these results one can conclude that the optimal number of parameters we should use for ARIMA modeling are five MA parameters and a constant as well as one seasonal parameter.

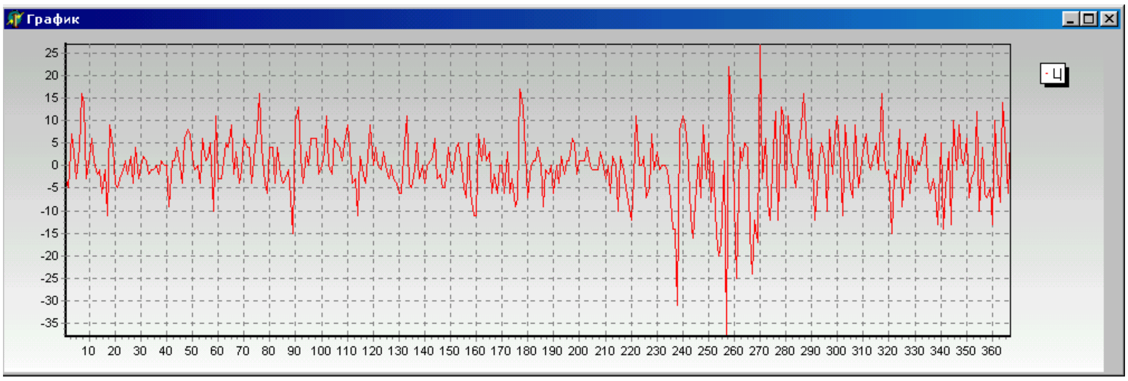

Figure 2. Stationary data obtained from Series B IBM

Table 1. Output from BSC criteria

(AR,MA)	-2.826(1,0)	-2.810(2,0)	-2.797(3,0)	-2.782(4,0)	-3.105(5,0)
-2.826(0,1)	-2.847(1,1)	-2.810(2,1)	-2.801(3,1)	-2.827(4,1)	-3.093(5,1)
-2.810(0,2)	-2.887(1,2)	-2.835(2,2)	-2.841(3,2)	-2.783(4,2)	-3.078(5,2)
-2.854(0,3)	-2.985(1,3)	-2.912(2,3)	-2.850(3,3)	-2.831(4,3)	-3.120(5,3)
-2.840(0,4)	-2.874(1,4)	-2.864(2,4)	-2.825(3,4)	-2.800(4,4)	-3.100(5,4)
-3.371(0,5)	-3.331(1,5)	-3.342(2,5)	-3.329(3,5)	-3.173(4,5)	-3.106(5,5)
-3.247(0,6)	-3.326(1,6)	-3.319(2,6)	-3.309(3,6)	-3.014(4,6)	-3.229(5,6)
-3.345(0,7)	-3.320(1,7)	-3.302(2,7)	-3.285(3,7)	-2.970(4,7)	-3.032(5,7)
-3.328(0,8)	-3.312(1,8)	-3.292(2,8)	-3.277(3,8)	-2.932(4,8)	-2.979(5,8)
-3.310(0,9)	-3.134(1,9)	-2.866(2,9)	-2.934(3,9)	-2.895(4,9)	-2.967(5,9)
-2.787(0,10)	-2.765(1,10)	-2.723(2,10)	-2.809(3,10)	-3.130(4,10)	-3.164(5,10)

The results on time series identification, in graphical and analytical forms, are depicted in Figure 2. We can see that this model fits data less than 50% of the time series movement. If we take '*R sqr-adj*' output, we can see that its value is 0.452 and it shows that this model can represent 45.2% of the time series or, in other words, it represents 45.2% better results than taking mean value of the series as a model. This can also be observed from the graphical output of the model in Figure 2. From the values of the Ljung-Box statistics we can see that residuals of the model represent the process of white noise.

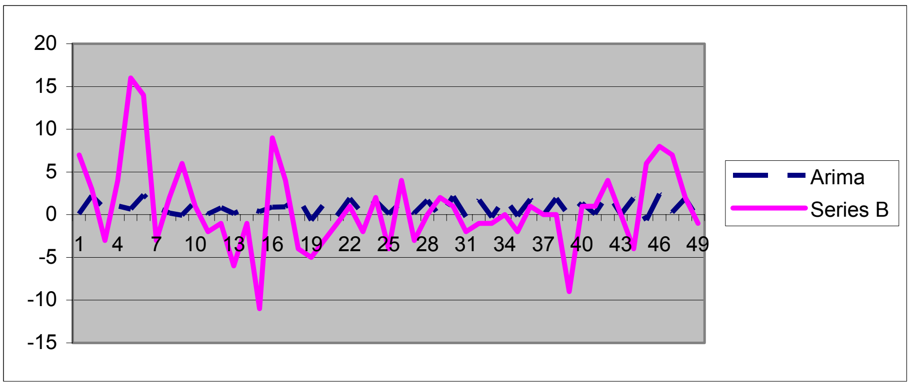

Figure 3. Series-graph and analytical output of the time series modelling with ARIMA model

```
Final Estimates of Parameters

Type            Coef      SE Coef         T        P
MA    1      -0.0604       0.0531     -1.14    0.257
MA    2      -0.0140       0.0533     -0.26    0.792
MA    3       0.0187       0.0532      0.35    0.726
MA    4       0.0329       0.0532      0.62    0.537
MA    5       0.0603       0.0536      1.12    0.261
SMA   6       0.9658       0.0192     50.34    0.000
Constant   -0.03592      0.01970     -1.82    0.069

Differencing:                 1 regular, 1 seasonal of order 6
Number of observations:       Original series 369, after differencing 362
Residuals:                    SS =  19125.4  (backforecasts excluded)
                              MS =     53.9  DF = 355

Modified Box-Pierce (Ljung-Box) Chi-Square statistic
Lag                   12        24        36        48
Chi-Square           8.0      29.8      41.6      59.8
DF                     5        17        29        41
P-Value            0.159     0.028     0.061     0.029
```

The results of the output obtained by ANN modelling with Eleman NN model are given in Figure 3. These are presented in the same format as the results from ARIMA model in series-graph and analytical data. From the output data we can determine the goodness of the fit. The value of '*R sqr-adj'* parameter is 0.945, and the sum of squared errors is 0.6768. We can conclude that this model fits 94.5% to the process dynamics of the given time series. We have implemented only one AR parameter in this model. By increasing the number of the dependent variables we can increase the fit goodness to the number of 4 variables. However, when further increase of the number of dependent variables is applied the results get worsened, hence the increase on the number of dependent variables do not contribute to significant improvement. When two AR parameters were included, 1% better results have been obtained than with one AR parameter. This is due to the recurrent mechanism of the ANN employed. This neural network takes into account previous values of the time series through the recurrent loops it possesses. Following the analyzed data and results, we can conclude that this time series could be better represented by a non-linear model rather than with the linear one, hence NARX is apparently advantageous.

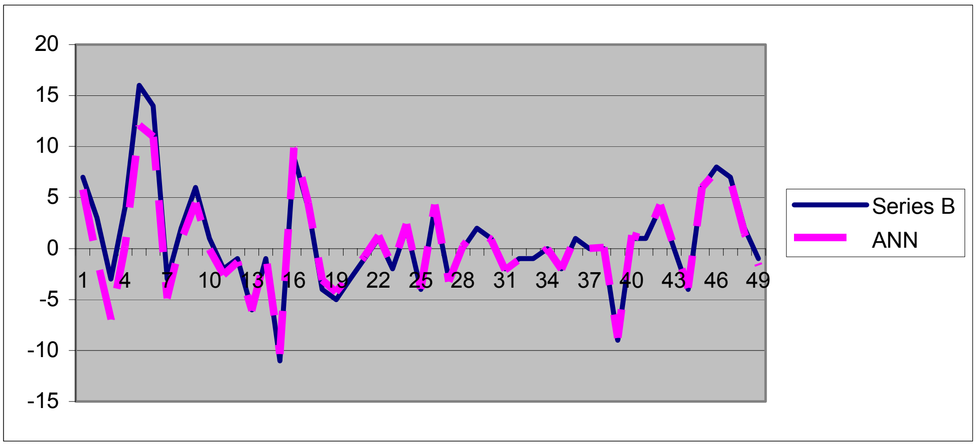

Figure 4. Graphical and analytical output of the time series modeling with ANN model

SSE=0.6768 R sqr=0.947 R sqr-adj=0.945

```
Modified Box-Pierce (Ljung-Box) Chi-Square statistic
Lag             6      12       24        36        48
Chi-Square   14.99   18.75    45.90     62.01     81.16
```

Should now the values of Ljung Box statistics on 5% level of significance are closely examined it may well be noticed the residuals obtained via this modelling represent a stationary time series. Namely, values at lags 6, 24, 36, and 48 are above those in the Chi-square distribution.

IV. CONCLUSION

Times series analysis and prediction has been successfully used to support the decision making in several real world application. Several different models can model economic processes and cycles. As we can see several techniques can be deployed for modeling and predicting time series, among which we highlight the Box-Jenkins approach and the Artificial Neural Networks (ANNs). While neural networks can fit a dataset much better than linear models it has been observed that they often forecast poorly which limits their appeal in applied settings [5].
However, both techniques for modeling and forecasting time series should be incorporated in educational process (if they are not implemented yet). They offer powerful tools for making analysis, which leads to understand movement of the observed time series as well as predict future values. Conclusions made with these models can be useful in planning and control future activities in many economic areas.

REFERENCES

[1] A.-M. S. Yaser, A. F. Atiya, M. Magdon-Ismail, and H. White, "Introduction to the special issue on neural networks in financial engineering". *IEEE Trans. on Neural Networks*, **NN-12** (4), pp. 653-656, July, 2001.
[2] G. E. P. Box, "Use and abuse of regression". *Technometrics*, **8**, pp. 625-629, 1966.
[3] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*. San Francisco CA: Holden-Day, 1970.
[4] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis: Forecasting and Control* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall, 1994.
[5] C. J. Andreeski, Non-Classical System-Theoretic Approach to Decision and Control in Actuary Assets Management (*PhD Thesis*), SS Cyril and Methodius University, Skopje, MK, 2003.
[6] G. Doffner, Neural Networks for Time Series Processing, Austrian Research Institute for AIR. Bastos, C. Predencio, T. B. Ludermir, Evolutionary design of NN, Recife Brasil
[7] X. Cheng, J. Racine, and N. R. Swanson, "Semiparametric ARX neural-network models with an application to forecasting inflation". *IEEE Trans. on Neural Networks*, **NN-12** (4), pp. 674-683, July, 2001.

APPENDIX: ADDITIONAL CASE-STDIES OF TIME SERIES *by STATISTICS / ARIMA* / and ELMAN ANN MODELS

Series A-1/ - U.S. HOG PRICE DATA ANNUAL: R sqr-adj=0.186

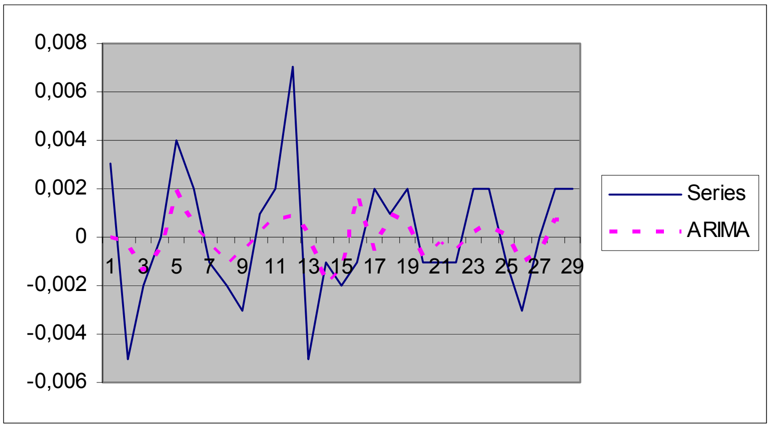

Figure A-1(a)

```
Final Estimates of Parameters
Type          Coef      SE Coef        T         P
MA    1     -0.1113      0.1066      -1.04     0.300
MA    2      0.1522      0.1087       1.40     0.165
MA    3      0.3774      0.1086       3.47     0.001

Number of observations:  81
Residuals:     SS =  0.000315245  (backforecasts excluded)
               MS = 0.000004042   DF = 78

Modified Box-Pierce (Ljung-Box) Chi-Square statistic
Lag                 12          24          36          48
Chi-Square         7.9        24.9        30.9        40.5
DF                   9          21          33          45
P-Value          0.542       0.251       0.572       0.663
```

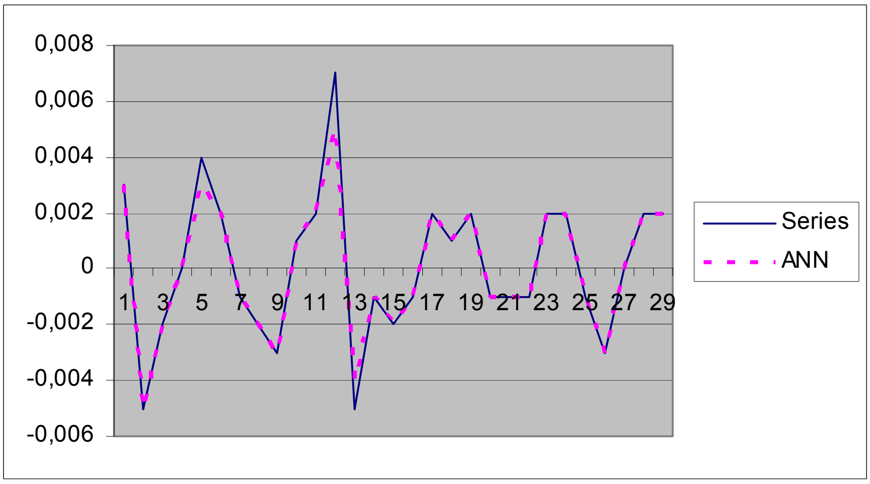

Figure A-1 (b)

SSE=0.1585 R sqr=0.980   R sqr-adj=0.974

```
Modified Box-Pierce (Ljung-Box) Chi-Square statistic
Lag                 12          24          36          48
Chi-Square         9.475       9.515       9.581       9.726
```

Series A-2/ - SUNSPOT NUMBERS YEARLY: R sqr-adj=0.373

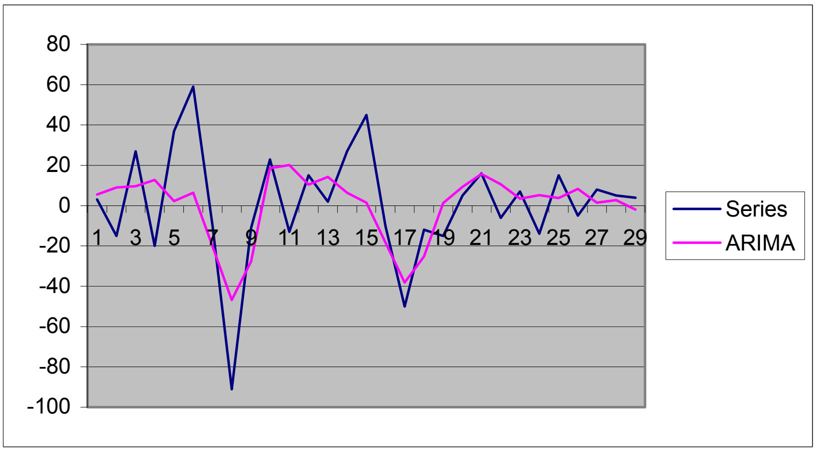

Figure A-2 (a)

```
Final Estimates of Parameters
Type            Coef      SE Coef         T        P
AR   1        0.8315       0.0911      9.13    0.000
AR   2       -0.4845       0.0916     -5.29    0.000
MA   1        0.9774       0.0175     55.82    0.000

Number of observations:   98
Residuals:     SS =   26974.8   (back-forecasts excluded)
               MS =    283.9   DF = 95

Modified Box-Pierce (Ljung-Box) Chi-Square statistic
Lag                   12        24        36        48
Chi-Square          26.2      34.0      38.9      42.3
DF                     9        21        33        45
P-Value            0.002     0.036     0.223     0.587
```
[2]

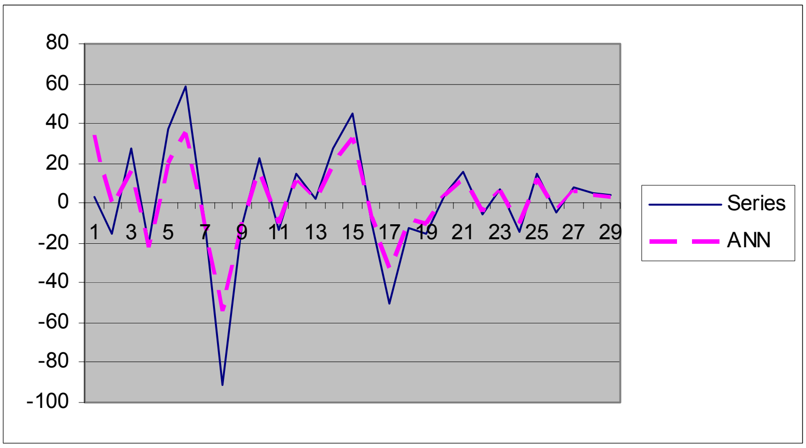

Figure A-2 (b)

SSE=0.6981 R sqr=0.866   R sqr-adj=0.835

[2] The results of ARIMA model are obtained by means of MINITAB ver. 13

```
Modified Box-Pierce (Ljung-Box) Chi-Square statistic
Lag                  12          24          36          48
Chi-Square        33.17       43.24       45.01       46.43
```

Series A-3/ – CHEMICAL PROCESS TEMPERATURE READINGS MINUTELY: R sqr-adj=0.0841

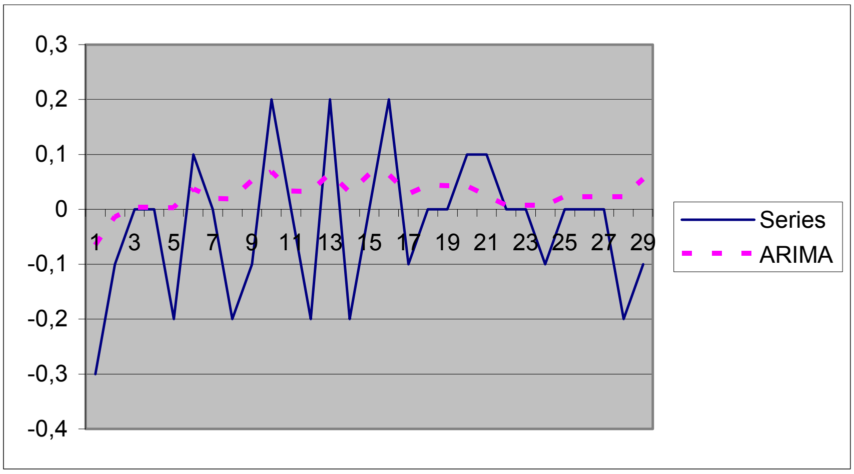

Figure A-3 (a)

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
AR 1	0.8116	0.0411	19.75	0.000
MA 1	0.9804	0.0017	589.11	0.000

Number of observations: 224
Residuals: SS = 4.06478 (backforecasts excluded)
MS = 0.01831 DF = 222

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	12.6	27.1	50.0	54.8
DF	10	22	34	46
P-Value	0.247	0.207	0.038	0.176

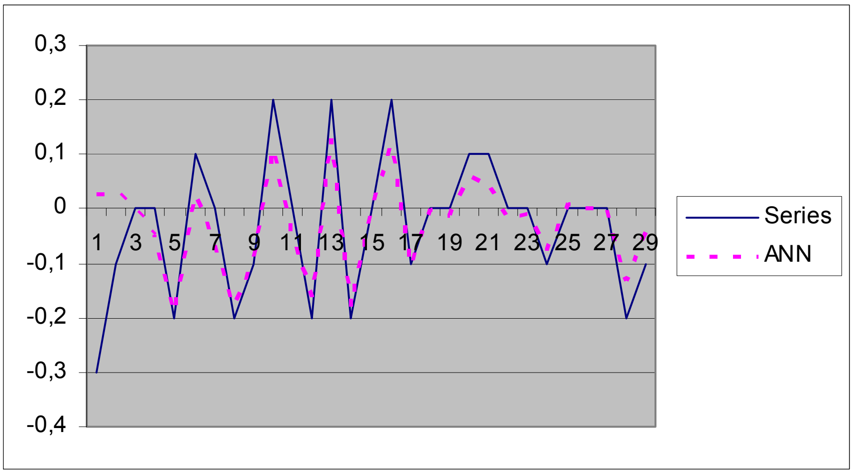

Figure A-3 (b)

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	22.74	31.24	56.49	66.49

Series A-4/ - CHEMICAL PROCESS CONCENTRATION READINGS MINUTELY:
R sqr-adj=0.281

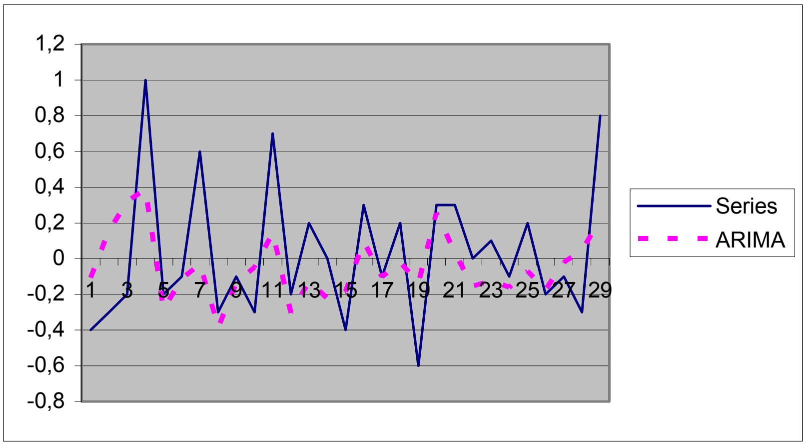

Figure A-4 (a)

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
AR 1	0.2187	0.0940	2.33	0.021
MA 1	0.8253	0.0539	15.31	0.000

Number of observations: 196
Residuals: SS = 19.2792 (backforecasts excluded)
MS = 0.0994 DF = 194

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	11.3	25.1	46.7	51.1
DF	10	22	34	46
P-Value	0.335	0.294	0.071	0.281

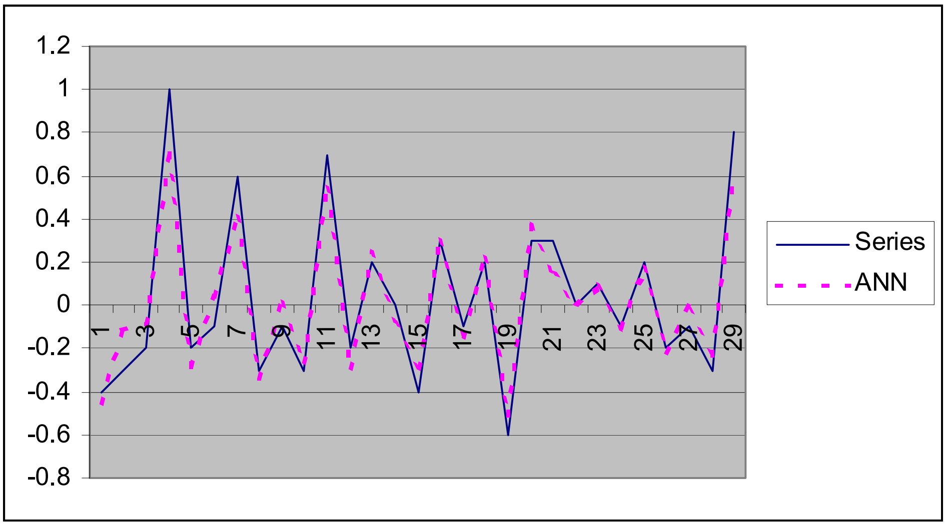

Figure A-4 (b)

SSE=1.2197 R sqr=0.911 R sqr-adj=0.902

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	40.61	62.71	81.60	85.31