A Comparison of ARIMA, Neural Network and Linear Regression Models for the Prediction of Infant Mortality Rate

Purwanto [1], Chikkannan Eswaran [2] , Rajasvaran Logeswaran [3]
[1] Faculty of Computer Science, Dian Nuswantoro University
50131 Semarang, Indonesia
[1, 2] Faculty of Information Technology, Multimedia University
63100 Cyberjaya, Malaysia
[3] Faculty of Engineering, Multimedia University
63100 Cyberjaya, Malaysia
[1] mypoenk@gmail.com, [2] eswaran@mmu.edu.my, [3] loges@mmu.edu.my

***Abstract*—The aim of this paper is to compare the performances of ARIMA, Neural Network and Linear Regression models for the prediction of Infant Mortality Rate. The performance comparison is based on the Infant Mortality Rate data collected in Indonesia during the years 1995 – 2008. We compare the models using performance measures such as Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE). The results show that the Neural Network model with 6 input neurons, 10 hidden layer neurons and using hyperbolic tangent activation functions for the hidden and output layers is the best among the different models considered.**

Keywords: ARIMA, Neural Network, Linear Regression, Infant Mortality Rate, Mean Absolute Error, Mean Absolute Percentage Error, Root Mean Square Error

I. INTRODUCTION

The infant mortality Rate (IMR) in a community is a widely used indicator of general health status. The IMR reflects a broad range of social, economic and medical conditions. Communities with problems such as unemployment, poverty, and low literacy tend to have higher infant mortality rates [1]. In Indonesia, the mortality statistic such as infant mortality is regarded as one of the important instruments to monitor and evaluate national health development.

The time series forecasting is employed in various applications to predict the future value using the past time series data. Many models have been applied in time series forecasting such as ARIMA, Logistic Regression and Neural Network. Choosing the best model is needed in the time series forecasting. These models have been used for different applications and from the result reported, it is clear that the performances of these models vary depending on the type of data used. In this paper, we present the performance comparison of these models for specific application, namely, prediction of IMR which is very useful for healthcare management. From this point of view, the result reported in this paper may be considered to be a significant contribution in the area of healthcare management.

The prediction or forecasting of IMR assumes importance for health departments since a good and accurate prediction is very helpful in devising appropriate action plans.

II. PREDICTION MODELS USED

This section describes the theory of the various models implemented in this paper.

A. ARIMA Model

Time series analysis has several objectives: modeling, forecasting and controlling. Forecasting deals with the issue of constructing models and methods that can be used to produce accurate short-term predictions [2]. The aim of modeling is to build a statistical model that adequately represents the long-term behavior of a time series. These goals are not necessarily identical. While the former frequently leads to a black box model that produces predictions, the objective of the latter is more towards finding the model that has generated the data.

A univariate time series $\{X_t\}$ is a series of observations of a variable over discrete intervals of time. These observations are equally spaced in time.

$$\{X_t\} = (x_1, x_2, x_3, ..., x_t) \tag{1}$$

The time series analysis involves modeling the series as a function of its past observations and *errors* (*residuals*). Error (e_t) is the difference between the observed value (x_t) and the forecast value ($\hat{x}_t$):

$$e_t = x_t - \hat{x}_t \tag{2}$$

The two fundamental building blocks of a linear univariate time series model are the autoregressive (AR) model and moving average (MA) model. The forecast in AR

978-0-7695-4062-7/10 $26.00 © 2010 IEEE
DOI 10.1109/AMS.2010.20

model is a function of its past observations, while in a MA model the forecast is a function of its past errors.

An autoregressive model of order p, AR(p) is given by

$$\hat{x}_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + ... + \phi_p x_{t-p} \tag{3}$$

A moving average model of order q, MA(q) is given by

$$\hat{x}_t = \theta_1 e_{t-1} + \theta_2 e_{t-2} + ... + \theta_q e_{t-q} \tag{4}$$

The *order* of the model is defined by the highest term present in the describing equation. The order of an AR polynomial is denoted by p and that of a MA polynomial is denoted by q.

An autoregressive moving average (ARMA) process is a combination of autoregressive and moving average polynomials in a single equation. An ARMA (p, q) model is defined by [3]:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + ... + \phi_p x_{t-p} + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + ... + \theta_q e_{t-q} \tag{5}$$

where the orders of AR and MA polynomials are p and q respectively.

In order to make the notational form simple, a special operator called the backshift operator (B) is used. The backshift operator is defined by

$$B^j x_t = x_{t-j} \tag{6}$$

where j = 0, 1, 2, ...

Finally, an ARMA (p, q) model can be expressed as

$$\phi(B) x_t = \theta(B) e_t \tag{7}$$

where $\phi(.)$ and $\theta(.)$ are the p^{th} and q^{th} degree polynomials are given by [4]:

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - ... - \phi_p z^p$$

and (8)

$$\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + ... + \theta_q z^q$$

ARMA models employed for time series data using ordinary differencing are called autoregressive integrated moving average (ARIMA) models. In an ARIMA (p, d, q) model, the term d denotes the order of differencing. An ARIMA model with order of differencing d can be expressed as [4]:

$$\phi(B)(1-B)^d x_t = \theta(B) e_t \tag{9}$$

B. Neural Network Model

Neural networks (NN) for forecasting have been investigated by many researchers over past several years. The motivation for using neural networks is based on the fact that these models are capable of handling non-linear relationships. Examples of neural networks used for time series forecasting purposes can be found in [5, 6, 7, 8].

Neural Network model consists of different layers which are connected to each other by connection weights. Between the extremities of the input layer and the output layer are the hidden layers. The nodes in each layer are connected by flexible weights, which are adjusted based on the error or bias.

The function of the input layer is for data entry, data processing takes place in the hidden layer and the output layer functions as the data output result. Fig. 1 shows the architecture of a general multilayer backpropagation neural network.

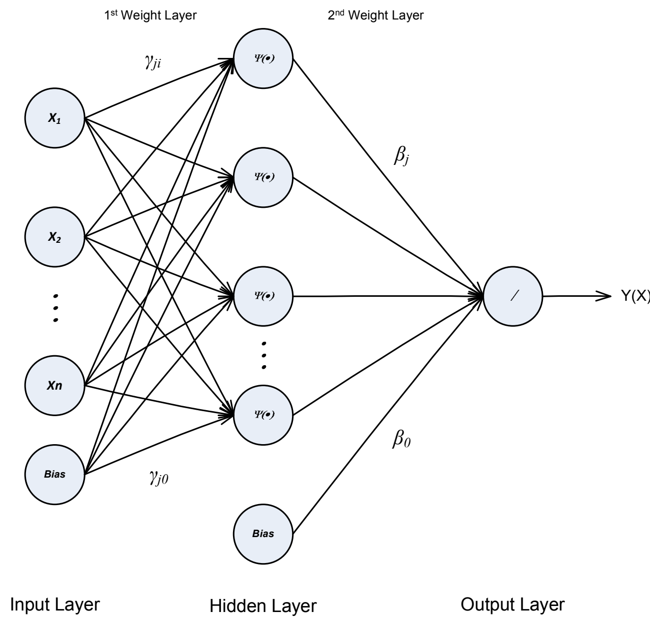

Figure 1. A general Multilayer Backpropagation Neural Network

In this architecture, the response value $Y(x)$ is computed as [5]:

$$Y(x) = \beta_0 + \sum_{j=1}^{H} \beta_j \psi(\gamma_{j0} + \sum_{i=1}^{n} \gamma_{ji} x_i) \tag{10}$$

where (β_0, β_1, …, β_H, γ_{10},…, γ_{Hn}) are the weights or parameters of the Neural Network. The non linearity enters into the function $Y(x)$ through the so called activation function ψ.

Information processing in every neuron is done by summing the multiplication result of connection weights with input data. The result is transferred to the next neuron through the activation function. There are several kinds of activation functions ψ, such as sigmoid, bipolar sigmoid and hyperbolic tangent.

The Hyperbolic Tangent activation function [9] given below:

$$\psi(x_j) = \tanh(x_j) = \frac{e^{x_j} - e^{-x_j}}{e^{x_j} + e^{-x_j}} \tag{11}$$

C. Linear Regression

The regression model describes the mean of the normally distributed dependent variable Y as a function of the predictor or independent variable x [10]:

$$Y_i = \beta_0 + \beta_1 \, x_i + \varepsilon_i \tag{12}$$

where Y_i is the value of the response or dependent variable from the i-th pair, β_0 and β_1 are the two unknown parameters, x_i is the value of the independent variable from the i-th pair, and ε_i is a random error term.
The predicted or estimated or fitted values of the regression model are calculated as [11]:

$$\hat{Y}_i = b_0 + b_1 x_i \tag{13}$$

The parameters b_0 and b_1 in the above equation are computed as:

$b_0 = \mu - \beta_1 (n+1)/2$

$b_1 = [\theta - n \mu (n+1)/2] / [\sigma^2 - n (n+1)^2 /4]$

where

$$\mu = \sum_{i=1}^{n} x_i / n$$

$$\theta = \sum_{i=1}^{n} i.x_i \tag{14}$$

$$\sigma^2 = \sum_{i=1}^{n} x_i^2$$

This model is referred to as the linear regression (LR) model [11]. It is linear because the independent variable appears only in the first power; if we plot the mean of Y versus x, the graph will be a straight line with intercept b_0 and slope b_1.

D. Performance Measures

The prediction models are evaluated in terms of their ability to forecast the future values. Several measures are used in comparing forecasting performance of different models. The most common measure is the Root Mean Square Error (RMSE). The other measures that are used are the mean absolute error (MAE) and the mean absolute percentage error (MAPE).

The Root Mean Square Error and mean absolute percentage error are used for comparison of model accuracy. Lower values are better. The MAE measures the average magnitude of the error. The RMSE is likely to be used for data that has the undesirable large error. And both MAE and RMSE can be used together to diagnose the variation in the errors in a set of forecast. If value of RMSE > MAE then there is a variation in the errors. They are negatively-oriented scores. Lower values are better.
The three forecast error statistics are computed as follows:

1) Mean Absolute Error (MAE)

The MAE measures the average magnitude of the errors in a set of forecasts, without considering their direction. [12]:

$$MAE = \frac{\sum_{t=1}^{n} \left| Y_t - \hat{Y}_t \right|}{n} \tag{15}$$

2) Mean Absolute Percentage Error (MAPE)

MAPE produces a measure of relative overall fit [12]:

$$MAPE = \frac{\sum_{t=1}^{n} \frac{\left| Y_t - \hat{Y}_t \right|}{Y_t} * 100}{n} \tag{16}$$

3) Root Mean Square Error (RMSE)

The RMSE is a quadratic scoring rule which measures the average magnitude of the error. [12, 13]

$$RMSE = \frac{\sum_{t=1}^{n} \left(Y_t - \hat{Y}_t \right)^2}{n} \tag{17}$$

III. DATA

The infant mortality rate data used in this study are collected in Indonesia from Indonesian Health Profile, Ministry of Health Republic of Indonesia. Table 1 shows the IMR data during 1995-2008.

Table 1. IMR in Indonesia during 1995-2008

NO	YEAR	IMR
1	1995	55.00
2	1996	54.00
3	1997	52.00
4	1998	50.00
5	1999	44.00
6	2000	47.00
7	2001	50.00
8	2002	35.00
9	2003	38.00
10	2004	36.00
11	2005	36.00
12	2006	34.00
13	2007	32.00
14	2008	31.00

The statistical features of the IMR data such as the minimum, maximum, mean, standard deviation and Variance are shown in Table 2.

Table 2. . Descriptive Statistics of IMR in Indonesia

Name	Min	Max	Mean	Std. Dev.	Variance
IMR	31.00	55.00	42.4286	8.72410	76.110

IV. RESULTS

In this section, we compare the three prediction models. The models are based on ARIMA, Neural Network and Linear Regression. The results achieved by each implemented model are discussed below.

A. *ARIMA Model*

We compute prediction using ARIMA models with different parameter (*p, d, q*) values. The result of the time series forecasting with different ARIMA models is shown in Table 3.

Table 3. Result of the time series forecasting with ARIMA

Year	Actual Data	ARIMA (1,0,0)	ARIMA (0,0,1)	ARIMA (1,0,1)	ARIMA (0,1,1)	ARIMA (1,1,0)	ARIMA (1,1,1)	ARIMA (0,1,0)
1995	55	42.62	42.42	42.94				
1996	54	53.40	48.14	53.60	53.00	53.11	53.01	53.15
1997	52	52.53	45.78	53.10	51.49	51.72	51.44	52.15
1998	50	50.79	46.25	51.54	49.66	50.16	49.66	50.15
1999	44	49.05	44.79	49.78	47.74	48.16	47.76	48.15
2000	47	43.82	41.92	45.06	44.99	43.92	45.64	42.15
2001	50	46.43	45.68	46.32	43.32	42.96	43.07	45.15
2002	35	49.05	45.20	48.75	42.27	45.96	41.29	48.15
2003	38	35.98	35.87	38.29	39.36	38.89	40.63	33.15
2004	36	38.60	43.79	38.42	37.20	33.96	37.52	36.15
2005	36	36.85	37.41	36.99	35.08	34.16	35.38	34.15
2006	34	36.85	41.51	36.71	33.16	33.28	33.12	34.15
2007	32	35.11	37.59	35.20	31.22	32.16	31.21	32.15
2008	31	33.37	38.83	33.44	29.28	30.16	29.29	30.15
2009		32.50	37.39	32.37	27.40	28.72	27.26	29.15
MAE		3.853	5.790	3.737	2.183	2.543	2.184	2.781
MAPE		8.452	13.591	9.135	5.412	6.334	5.413	7.013
RMSE		6.022	7.133	6.176	3.375	4.337	3.450	4.705

From Table 3, it is noted that the smallest values of MAE, MAPE and RMSE are obtained for the model ARIMA(0,1,1). So, we can conclude that ARIMA(0,1,1) is the best ARIMA model.

Fig. 2 illustrates the actual and predicted graph of ARIMA models, where the horizontal axis represents the time (year) and the vertical axis is the value of IMR.

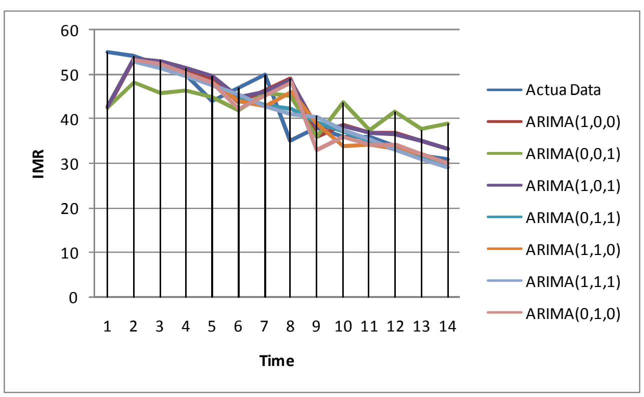

Figure 2. The actual and prediction graph of ARIMA Model

B. *Neural Network Model*

We also computed prediction using a multi-layer perceptron Neural Network model. Architecture configurations with different numbers of input and hidden layer neurons were tested to determine the optimum setup. Similarly, different activation functions such as hyperbolic tangent, bipolar sigmoid and sigmoid functions were tested. From the experimental results, it is found that the neural network model with 6 input neurons, 10 hidden layer neurons and using hyperbolic tangent activation functions for the hidden and output layers yields the minimum values for MAE, MAPE and RMSE.

Table 4. Result of the time series forecasting with Neural Network

Year	Actual Data	FORECASTING NO. OF INPUT NEURONS					
		6	5	4	3	2	1
1995	55						
1996	54						50.10
1997	52					52.24	49.63
1998	50				47.87	48.77	48.58
1999	44			43.62	47.04	43.22	47.36
2000	47		47.19	47.54	45.73	47.38	42.64
2001	50	49.53	49.66	49.52	50.22	46.94	45.20
2002	35	35.28	35.17	35.55	35.01	38.85	47.36
2003	38	38.04	37.90	37.86	38.43	38.72	34.50
2004	36	36.20	36.11	36.33	35.67	37.66	36.95
2005	36	36.09	35.97	35.57	34.87	34.41	35.26
2006	34	34.17	33.96	32.08	33.91	33.89	35.26
2007	32	32.02	31.41	32.64	32.93	31.63	33.82
2008	31	30.42	31.62	32.30	31.57	30.53	32.69
2009		31.73	31.48	32.45	31.11	30.25	32.23
MAE		0.234	0.242	0.672	0.923	1.204	3.271
MAPE		0.395	0.474	1.453	1.560	2.583	6.967
RMSE		0.300	0.322	0.842	1.287	1.649	4.389

The results of the time series forecasting using the optimum Neural Network model are shown in Table 4. For the sake of comparison, the results obtained for the different number of input neurons are also shown in Table 4.

Fig 3. Illustrates the actual and predicted graph of the optimum Neural Network models, where the horizontal axis represents the time (year) and the vertical axis is the value of IMR.

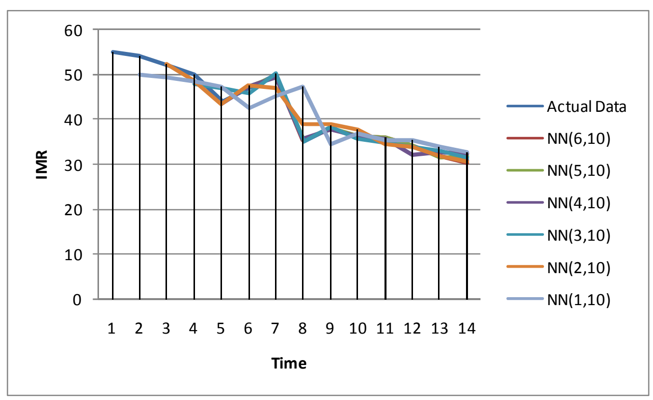

Figure 3. The actual and prediction graph of Neural Network Model

C. *Linear Regression*

The Linear Regression model for time series forecasting is obtained as follows:

$$Y_i = 57.198 - 1.969\ i \tag{18}$$

where Y_i is the predicted value for i-th year.

The result of the time series forecasting with Linear Regression model is shown in Table 5.

Table 5. Result of the time series forecasting with Linear Regression

Year	Actual Data	FORECASTING
1995	55	55.23
1996	54	53.26
1997	52	51.29
1998	50	49.32
1999	44	47.35
2000	47	45.38
2001	50	43.41
2002	35	41.44
2003	38	39.47
2004	36	37.51
2005	36	35.54
2006	34	33.57
2007	32	31.60
2008	31	29.63
2009		27.66
MAE		1.858
MAPE		4.533
RMSE		2.767

Fig 4. Illustrates the actual and predicted graph of Linear Regression model where the horizontal axis represents the time (year) and the vertical axis is the value of IMR.

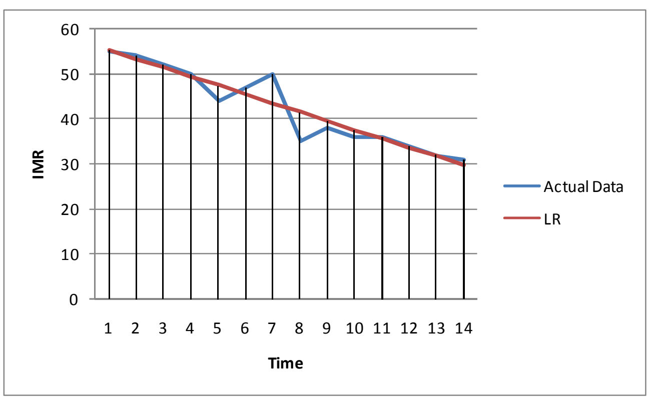

Figure 4. The actual and prediction graph of Linear Regression Model

D. *Comparison of Models*

Table 6 shows the comparison of MAE, MAPE and RMSE obtained with ARIMA, Neural Network and Linear Regression models.

From Table 6, it is seen that the best model for the prediction of Infant Mortality Rate is the Neural Network Model with 6 input neurons, 10 hidden layer neurons (NN(6, 10)) using hyperbolic tangent activation functions

Table 6. Performance Measure of Models

MODELS	PERFORMANCE MEASURE		
	MAE	MAPE	RMSE
ARIMA(1,0,0)	3.853	8.452	6.022
ARIMA(0,0,1)	5.790	13.591	7.133
ARIMA(1,0,1)	3.737	9.135	6.176
ARIMA(0,1,1)	2.183	5.412	3.375
ARIMA(1,1,0)	2.543	6.334	4.337
ARIMA(1,1,1)	2.184	5.413	3.450
ARIMA(0,1,0)	2.781	7.013	4.705
NN(6,10)	0.234	0.395	0.300
NN(5,10)	0.242	0.474	0.322
NN(4,10)	0.672	1.453	0.842
NN(3,10)	0.923	1.560	1.287
NN(2,10)	1.204	2.583	1.649
NN(1,10)	3.271	6.967	4.389
LR	1.858	4.533	2.767

Fig.5 compares the performance measures of different models graphically. The horizontal axis represents the models and the vertical axis is the values of performance measure.

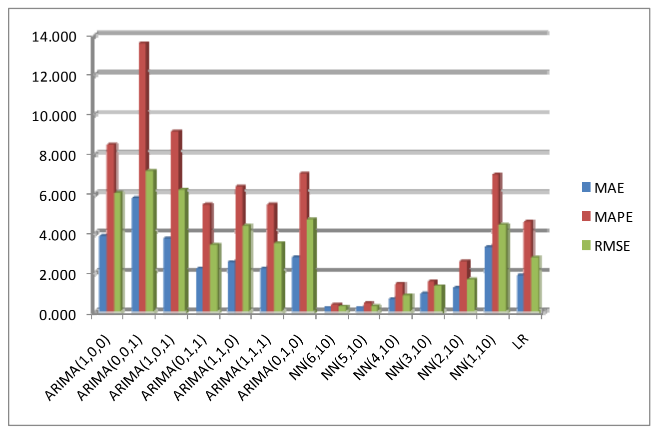

Figure 5. Comparison performance measure of models

V. CONCLUSION

This paper has described three prediction models for Infant Mortality Rate. The prediction was done using ARIMA, Neural Network and Linear Regression models. The performances of the models are compared using the data of Infant Mortality Rate collected during 1995 – 2008 in Indonesia. Performance measures such as Mean Absolute Error, Mean Absolute Percentage Error and Root Mean Square Error have been employed for comparison. From the results, it was found that the Neural Network model with 6 input neurons, 10 neurons in the hidden layer and using hyperbolic tangent activation functions for the hidden and output layers yields the best forecasting result.

REFERENCES

[1] Condran, Gretchen., Murphy, Jennifer. "The Infant Mortality Rate in Late Nineteenth and Early Twentieth Century Philadelphia" , The annual meeting of the American Sociological Association, Philadelphia, PA, 2005.

[2] Chatfield, C., "Time Series Forecasting", Chapman & Hall, London, 2001.

[3] Andre´ Luis S. Maia, Francisco de A.T. de Carvalho, Teresa B. Ludermir, "Forecasting models for interval-valued time series", Neurocomputing Vol.71, 2008, pp 3344–3352.

[4] Peter J. Brockwell Richard A. Davis, "Introduction to Time Series and Forecasting", Springer-Verlag New York, Inc, 2002.

[5] Suhartono, "*Feedforward Neural Networks Untuk Pemodelan Runtun Waktu"*, Gajah Mada University, Indonesia, 2008.

[6] I. Kaastra, M. Boyd , "Designing a neural network for forecasting financial and economic time-series", Neurocomputing Vol. 10, 1996, pp 215–236.

[7] H. R. Maier, G. C. Dandy, "Neural network models for forecasting univariate time series", Neural Networks World 6 (5), 1996, pp 747–772.

[8] R. Sharda, R. Patil, "Neural networks as forecasting exports: an empirical test, in: Proceedings of IJCNN Meeting Vol. 2, 1990, pp. 491–494.

[9] SPSS Inc, "SPSS 16.0 Algorithms", SPSS United States of America, 2007.

[10] Chap T. Le, "Introductory Biostatistics", A John Wiley & Sons Publication, 2003.

[11] Jamie DeCoster, "Notes on Applied Linear Regression", Free University Amsterdam, 2003.

[12] Luis A. Dı´az-Robles, Juan C. Ortega et al, "A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile", Atmospheric Environment 42, 2008.

[13] I. Rojas, O. Valenzuela, et.al, "Soft-computing techniques and ARMA model for time series prediction", Neurocomputing, 2008, pp 519 – 537.