Paper

# Performance Comparison of Four New ARIMA-ANN Prediction Models on Internet Traffic Data

C. Narendra Babu and B. Eswara Reddy

*Department of Computer Science and Engineering, JNT University College of Engineering, Anantapuramu, India*

**Abstract—Prediction of Internet traffic time series data (TSD) is a challenging research problem, owing to the complicated nature of TSD. In literature, many hybrids of auto-regressive integrated moving average (ARIMA) and artificial neural networks (ANN) models are devised for the TSD prediction. These hybrid models consider such TSD as a combination of linear and non-linear components, apply combination of ARIMA and ANN in some manner, to obtain the predictions. Out of the many available hybrid ARIMA-ANN models, this paper investigates as to which of them suits better for Internet traffic data. This suitability of hybrid ARIMA-ANN models is studied for both one-step ahead and multi-step ahead prediction cases. For the purpose of the study, Internet traffic data is sampled at every 30 and 60 minutes. Model performances are evaluated using the mean absolute error and mean square error measurement. For one-step ahead prediction, with a forecast horizon of 10 points and for three-step prediction, with a forecast horizon of 12 points, the moving average filter based hybrid ARIMA-ANN model gave better forecast accuracy than the other compared models.**

***Keywords—ANN, ANN training, ARIMA, Box-Jenkins methodology, hybrid ARIMA-ANN model, Internet traffic forecasting.***

## 1. Introduction

Time series data (TSD) forecasting has its applications in various domains like agricultural, climatic, econometric, financial and communication. With the growing telecommunication sector, the service providers should be able to effectively distribute their resources for continued services. Internet traffic data forecasting helps service providers manage available bandwidth and resources properly. Consider a situation, where a large part of the bandwidth is being used by a network. Within the next half an hour, if it is a priori known that this network will not consume more than 30% of the available bandwidth, the service provider can reduce the network bandwidth and in-turn divert the rest of the available bandwidth to some other network. This way the resources can be used optimally. Hence, prediction of Internet traffic TSD is drawing more attention in the present days.

## 2. Related Work

Autoregressive integrated moving average (ARIMA) linear models are popularized by Box and Jenkins in 1970 for time series prediction. These models are applied on various TSD such as electricity prices [1], [2], sugar prices [3], stock market data [4], and wind speed data [5], for the prediction of future values. Next, the pre-processing based ARIMA models were introduced. In [6], a wavelet transformation based ARIMA forecasting is done on global temperature data. In [7], classification and feature extraction techniques were proposed for electrocardiography data. These preprocessing steps help to obtain more accurate predictions.

Later the era has been conquered by the ANN (non-linear) models. ANN was able to model a wide range of TSD compared to ARIMA, as they are capable of modeling non linear variations. ANNs have been applied to electricity demand data [8], financial data [9] river flow data [10], and network data [11], for prediction. Compared to ARIMA, these TSD were accurately predicted with ANN. In [12], neural networks were used to predict earthquakes in Chile.

Instead of individual ARIMA or ANN, research progressed in the direction of combining the benefits of both ARIMA and ANN models to devise hybrid ARIMA-ANN models. Next, a hybrid ARIMA-ANN versions was proposed by Zhang [13], which gave good prediction accuracy compared to individual models, when applied to Wolf's sunspot data, Canadian lynx data, and exchange rate data for one-step ahead prediction. Next, a new hybrid ARIMA-ANN method was proposed by Khashei and Bijari [14], which gave better performance. The hybrid model devised by Zhang was used for electricity price forecasting in [15] and water quality TSD prediction in [16]. In [17], a multiplicative model was proposed (Li Wang *et al.*), which is in contrast to the additive model of Zhang. The results showed that it is no less in comparison to the application of additive Zhang model. In [18], a moving average filter based hybrid ARIMA-ANN model is proposed which first decomposes the data and then applies the apt model on each decomposition. This model is shown to outperform both Zhang, Khashei and Bijari models, when applied to sunspot data, electricity price data and stock price data, in both one-step ahead and multi-step ahead forecasting.

Many other prediction models are available in the literature. Some of them are based on support vector machines (SVM) [19], and some others on fuzzy logic [20]. Spectral techniques based on SVD were proposed in [15] and the references therein. Most of the forecasting problems showed that hybrid models are a better solution. However, if the hybrid model involves large number of decompositions, the prediction accuracy suffers. Hence, a hybrid model should contain limited number of individual models to retain the model simplicity and prediction accuracy.

In this paper, the Internet traffic TSD predictions for both one-step ahead and multi-step ahead cases are obtained using individual ARIMA, ANN models, hybrid ARIMA-ANN models of Zhang [13], Khashei and Bijari [14], Wang *et al.* [17], Babu and Reddy [18]. From obtained results the best model was identified.

The rest of the paper is organized as follows. In Sections 3–4, the ARIMA, ANN, and some existing hybrid ARIMA-ANN models are described. In Section 5, the results are discussed in four subsections, along with tables of performance measures and graphs of predicted values. Section 6 ends the paper with a conclusion.

# 3. ARIMA and ANN Prediction Models

Some of the hybrid ARIMA-ANN models available in the literature are outlined, with a brief description.

### 3.1. ARIMA

To model a TSD using ARIMA, a training data is provided. ARIMA modeling fits a linear equation to this data if it is stationary. If the training data is non-stationary, differencing is performed till it becomes stationary. The corresponding order of differencing is notated as $d$. The moving average (MA) model order $q$ and auto-regressive (AR) model order $p$ are determined from the decaying nature of autocorrelation function (ACF) plot and the partial ACF (PACF) plot respectively. Detailed correlation analysis for order determination is given in [21]. According to the modeling procedure, the present value of TSD, $y_t$ is considered as a weighted sum of past data points $y_{t-1}, y_{t-2}, \ldots, y_{t-p}$ and error values $e_t, e_{t-1}, e_{t-2}, \ldots, e_{t-q}$. The model is shown in Eq. (1), where $y_{t-k}$ is the TSD value at a delay of $k$ time points. The model assumes that the error series $e_t$ has a gaussian distribution.

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \ldots a_p y_{t-p} + e_t + b_1 e_{t-1} + \ldots b_q e_{t-q} \tag{1}$$

The model coefficients $a_1, \ldots, a_p, b_1, \ldots, b_q$ are estimated using Box-Jenkins method [21]. As non-linear likelihood estimation is complex, Gaussian maximum likelihood estimation (GMLE) approaches [22] are used in the estimation of the model coefficients. The model is then validated using diagnostic checks like Akakine Information Criterion (AIC) and Bayesian Information Criterion (BIC). Also normality test like Jarque-Bera test, check on residual autocorrelation plots to meet the confidence limits are also performed. Once the model is identified best according to these diagnostic tests, it is selected for application on the TSD. The selected model is used to predict future TSD values over the prediction horizon.

### 3.2. ANN

Unlike ARIMA, ANN is a non-linear modeling technique. The neural network model architecture comprises of neurons, similar to the brain's architecture. For example a three layer ANN, with the three layers called as input, hidden and output layers is shown in Fig. 1 [18]. Each layer comprises of one or more nodes. For TSD prediction problem,

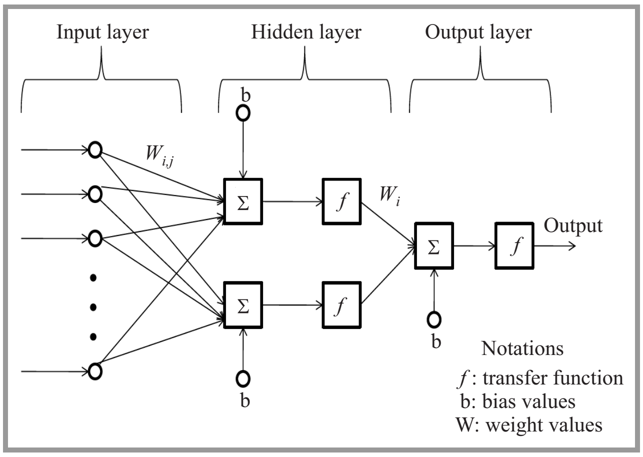


***Fig. 1.*** Three-layer ANN architecture.

the output layer has one node. The hidden layer can have any number of nodes, whose outputs are linked to the output node. The input layer can have one or more nodes depending on the number of TSD points involved in the prediction. There can be more than one hidden layer. The neurons are acyclically linked processing units. Three-layer ANNs are widely used for TSD forecasting. To model TSD using ANN, $y_t$ is expressed as a non-linear function $f$ of $y_{t-1}, \ldots, y_{t-A}$, where $A$ is the lag till which the TSD points are involved in prediction. The model equation is:

$$y_t = g(y_{t-1}, y_{t-2}, \ldots, y_{t-A}) + v_t, \tag{2}$$

where $v_t$ is the noise or error term. The transfer function of the hidden layer can be a linear, sigmoid, tan-sigmoid or log-sigmoid in nature. A sigmoid function is:

$$\mathrm{Sigmoid}(x) = \frac{1}{1+e^{-x}}. \tag{3}$$

The model coefficients in ANN are weights of each link and the corresponding bias values. To determine these values, a training data is provided to ANN. Many training algorithms are available [23], out of which Levenberg-Marquardt (LM) training algorithm is used in [14], [18]. In [13], a reduced gradient algorithm and in [16], a scaled conjugate gradient algorithm are used. Here LM training is incorporated. The model is diagnosed using validation and testing phase, where the mean square error convergence is verified. If the error is converging, the model is valid, else it is invalid. After the testing phase, the model is used in the prediction of future values.

# 4. Hybrid ARIMA-ANN Models

Often, the given data may have both linear and nonlinear characteristics. So, hybrid models using both ANN and ARIMA methods are better than individual models for obtaining accurate predictions. Four existing ARIMA-ANN hybrid models considered for discussion in this paper are illustrated as follows.

### *4.1. Zhang's Hybrid ARIMA-ANN Model*

In 2003, Zhang proposed a hybrid ARIMA-ANN model. It is based on the assumption that the given TSD is a sum of two components, linear and non-linear, given in:

$$y_t = L_t + N_t \,. \tag{4}$$

On the given TSD series $y_t$, ARIMA is fit and the linear predictions are obtained, $\hat{L}_t$, as:

$$\hat{L}_t = a_1 y_{t-1} + \ldots + a_p y_{t-p} + b_1 e_{t-1} + \ldots + b_q e_{t-q} + e_t \,. \tag{5}$$

The difference series is obtained by Eq. (6) on which ANN is fit and the predictions $\hat{N}_t$ are obtained using Eq. (7):

$$n_t = y_t - \hat{L}_t \,, \tag{6}$$

$$\hat{N}_t = f(n_t, n_{t-1}, \ldots, n_{t-A}) + v_t \,. \tag{7}$$

The hybrid model predictions are now obtained by summing the ARIMA and ANN predictions:

$$\hat{y}_t = \hat{L}_t + \hat{N}_t \,. \tag{8}$$

This model is suitable for both one-step ahead and multi-step ahead prediction. It is shown to be better than individual models in terms of prediction accuracy. The model is block diagram presented in Fig. 2.

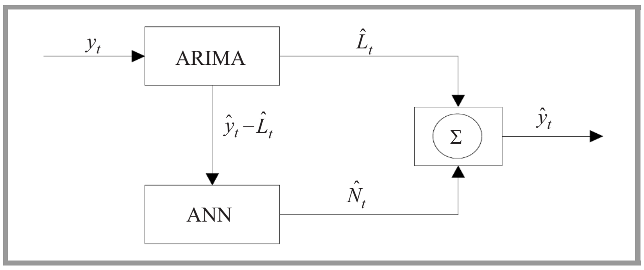


***Fig. 2.*** Zhang's hybrid ARIMA-ANN model.

### *4.2. Khashei and Bijari's Hybrid ARIMA-ANN Model*

In 2010, Khashei and Bijari proposed a new hybrid ARIMA-ANN model for TSD forecasting. Similar to Zhang's model, it also assumes that any TSD has linear and non-linear components, see Eq. (4). But the methodology adopted in prediction is different. An ARIMA model is fit on given TSD to obtain one forecast on the TSD using Eq. (5). Past original values, present prediction, and past error data are all input to the ANN. The ANN gets trained and once the model is validated, the one-step forecast of the given TSD is directly obtained from:

$$\hat{y}_t = f(\hat{L}_t, L_{t-1}, L_{t-2}, \ldots, L_{t-A}) + v_t \,. \tag{9}$$

It is shown to perform better than the Zhang's model in a variety of applications. It is suited for one-step forecasts, but for multi-step forecasting, the model is not suitable. If the past predictions are used as inputs instead of past original values, the model accuracy degrades. The model diagram is illustrated in Fig. 3.

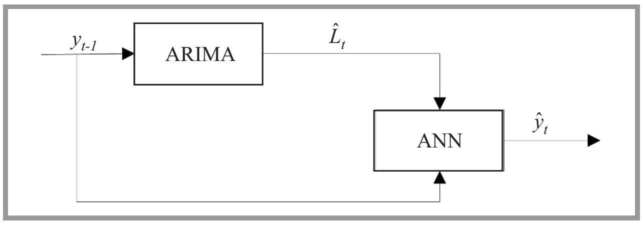


***Fig. 3.*** Khashei and Bijari's hybrid ARIMA-ANN model.

### *4.3. Multiplicative Hybrid ARIMA-ANN Model*

In 2013, Li Wang *et al*. proposed a multiplicative model for forecasting TSD, in contrast to the additive model proposed by Zhang. The model assumes that a given TSD is the product of a linear and a non-linear time series as:

$$y_t = L_t N_t \,. \tag{10}$$

The given TSD $y_t$ is modeled using ARIMA as shown in Eq. (5), similar to the same step in Zhang model. The predictions $\hat{L}_t$ obtained divide the original TSD to obtain the non-linear TSD series as:

$$n_t = \frac{y_t}{\hat{L}_t} \,. \tag{11}$$

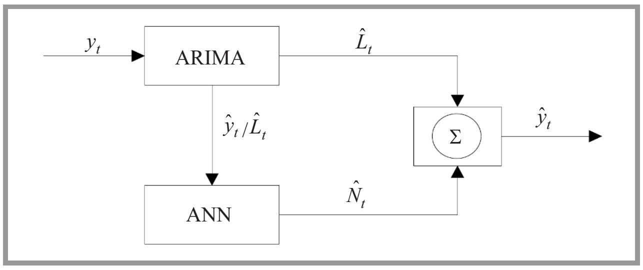


***Fig. 4.*** Multiplicative hybrid ARIMA-ANN model.

The series $n_t$ is modeled and predicted using ANN. The obtained non-linear predictions $\hat{N}_t$ in Eq. (6) and linear predictions $\hat{L}_t$ are multiplied to obtain the final model forecasts as given by Eq. (12). The block diagram of this model is as shown in Fig. 4.

$$\hat{y}_t = \hat{L}_t \hat{N}_t \,. \tag{12}$$

### *4.4. MA Filter Based Hybrid ARIMA-ANN Model*

In [18], a hybrid ARIMA-ANN model is devised using a decomposition step and then applying ARIMA and ANN suitably on each decomposition. The model framework assumes that any TSD is addition of a linear and a non-linear component given in Eq. (4) as in Zhang model. It also assumes that linear processes have less volatility compared to non-linear models, characterized by highly volatile nature. The steps of the model are:

1. An MA filter given by Eq. (13) is used to decompose the given TSD into a low volatile and a highly volatile component. The low volatile component is a smoothened TSD $y_{tr}$, and the highly volatile component is given by Eq. (14). The length of MA filter $m$ is adjusted such that one of the two decomposed time series is obtained with a kurtosis of 3, which is termed as low volatile decomposition $l_t$. The difference $h_t = y_t - l_t$ is considered highly volatile. The decomposition is indicated in Eq. (15).

$$y_{tr} = \frac{1}{m} \sum_{i=t-m+1}^{t} y_i \tag{13}$$

$$y_{res} = y_t - y_{tr} \tag{14}$$

$$y_t = l_t + h_t \tag{15}$$

2. The $l_t$ series is modeled and predicted using ARIMA model as in Eq. (16) to obtain $\hat{l}_t$. Note that this modeling using $l_{t-1}, l_{t-2}, \ldots, l_{t-p}$ unlike the ARIMA modeling step of Zhang (5), which uses $y_{t-1}, y_{t-2}, \ldots, y_{t-p}$.

$$\hat{l}_t = f(l_{t-1}, l_{t-2}, \ldots, l_{t-p}, e_t, e_{t-1}, \ldots, e_{t-q}) \tag{16}$$

3. The $h_t$ series is modeled and predicted using ANN model as shown in Eq. (17) to obtain $\hat{h}_t$.

$$\hat{h}_t = g(h_{t-1}, h_{t-2}, \ldots, h_{t-N}) + \varepsilon_t \tag{17}$$

4. The final model predictions are obtained by adding the predictions from steps 2 and 3:

$$\hat{y}_t = \hat{l}_t + \hat{h}_t \,. \tag{18}$$

The steps of this hybrid model are can be represented as a flow chart as shown in Fig. 5 [18].

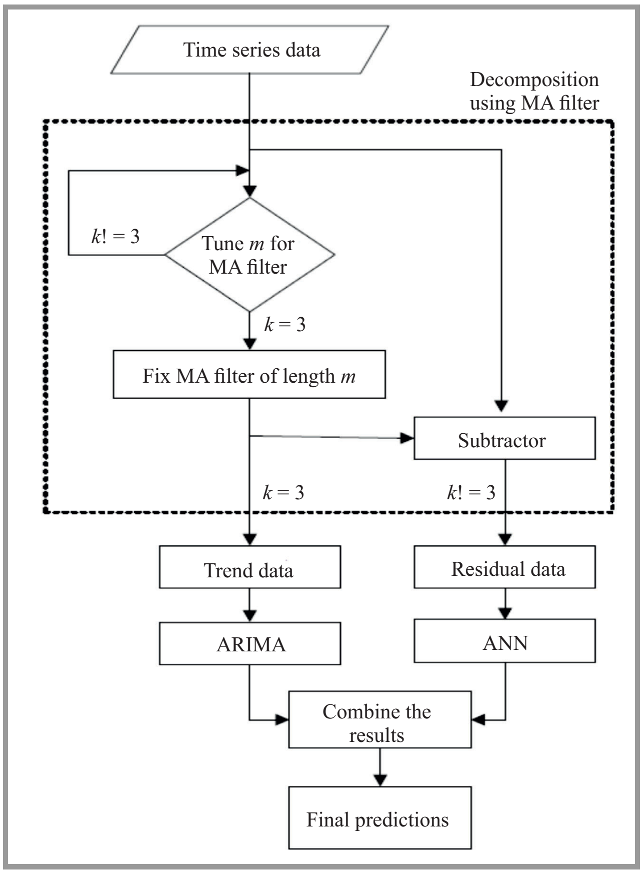


*Fig. 5.* MA filter based hybrid ARIMA-ANN model.

## 5. Results and Discussion

The prediction models ARIMA, ANN, Zhang's hybrid ARIMA-ANN, Khashei and Bijari's hybrid ARIMA-ANN, multiplicative hybrid ARIMA-ANN and MA-filter based hybrid ARIMA-ANN are extensively studied for their usage on Internet traffic data. The Internet traffic data obtained from [24] is used in this study. The raw data is

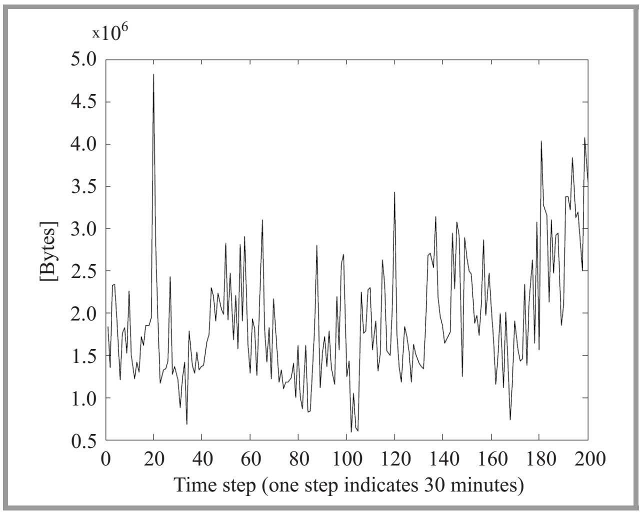


*Fig. 6.* Actual internet traffic TSD sampled at 30 min steps.

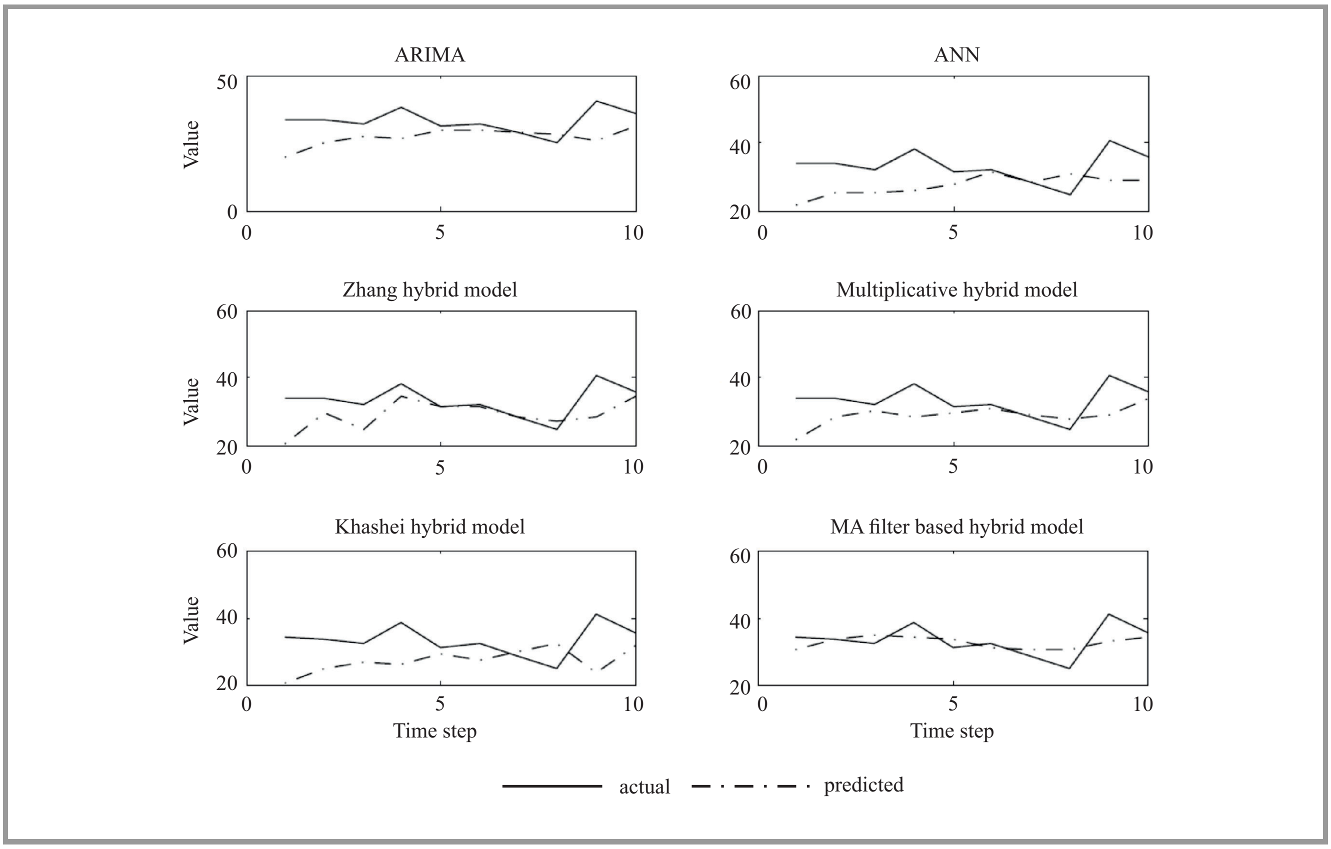


*Fig. 7.* One-step ahead predictions for TSD1.

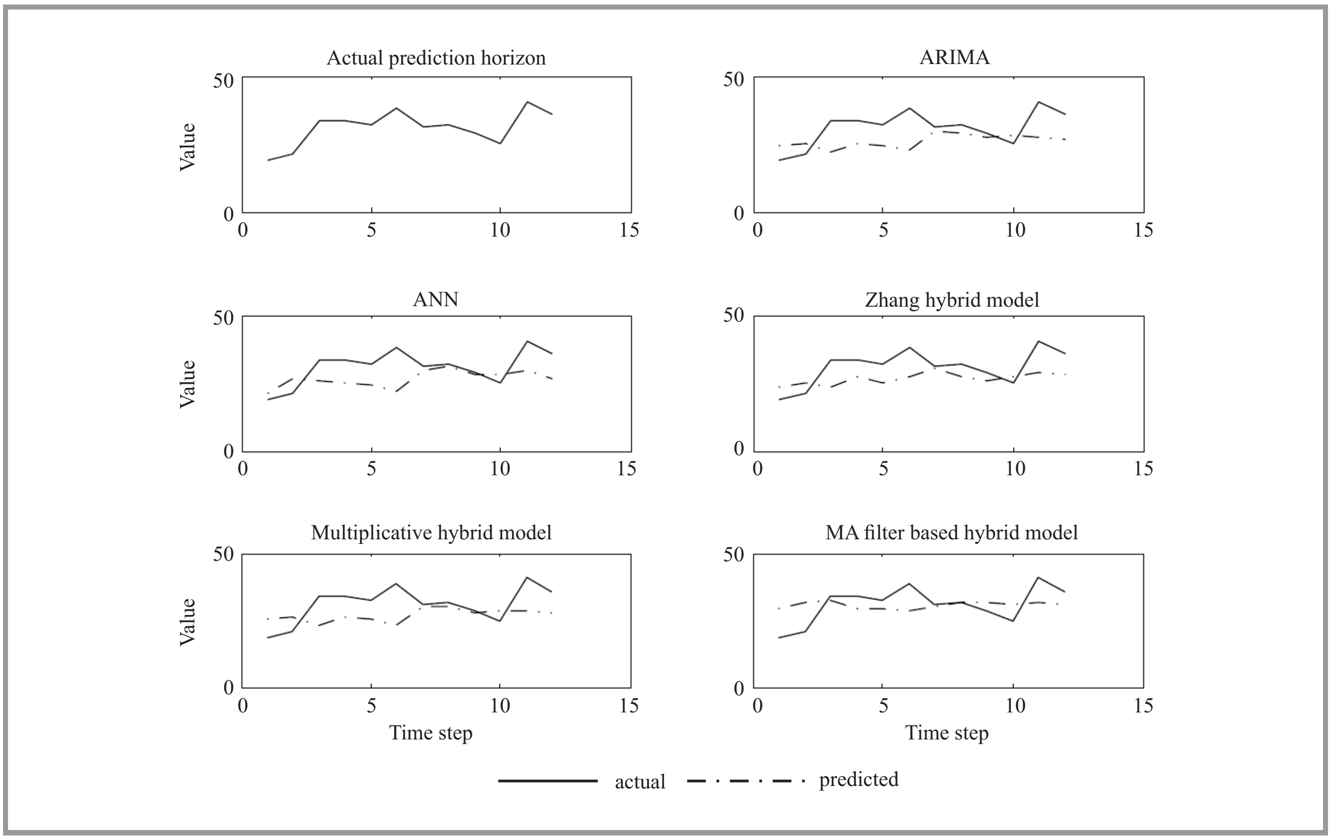


*Fig. 8.* Three-step ahead predictions for TSD1.

available at every one second for a period of 100 hr. This data is re-sampled to form two different data sets. The first TSD, named as TSD1 is obtained for every 30 min with a total number of 200 points. The second TSD, named as TSD2 is obtained for every 60 min with a total of 100 data points. The processed Internet data is in megabytes. To avoid big numbers, in this study, the authors divided this data by $10^5$ and then used it. However the raw data is plotted in Figs. 6 and 9 for 30 and 60 minutes sampling respectively. On both these Internet traffic datasets, the six models are applied and their performances are compared for both one-step ahead and three-step ahead forecasting. The performance measures considered in the study are described.

### 5.1. Performance Measurement

The two performance measures for accuracy comparison used in this paper are mean absolute error (MAE) and mean squared error (MSE), given by Eqs. (19) and (20) respectively. The smaller these values, the better is the model. In both formulas $E\{.\}$ is the expectation operation, $ni$ and $nf$ indicate start and end points of the prediction horizon, $y_{k,act}$ is the actual value of the time series, and $y_{k,pred}$ is the forecasted time series value at the instant $k$.

$$MAE = E\left\{\left|y_{act} - y_{pred}\right|\right\} = = \frac{1}{nf - ni + 1}\left(\sum_{k=ni}^{nf}\left|y_{k,act} - y_{k,pred}\right|\right) \tag{19}$$

$$MSE = E\left\{\left|y_{act} - y_{pred}\right|^2\right\} = \frac{1}{nf - ni + 1}\left(\sum_{k=ni}^{nf}\left|y_{k,act} - y_{k,pred}\right|^2\right) \tag{20}$$

### 5.2. Results for TSD1

TSD1 comprises of 200 points, each indicating the number of packets transmitted. The forecast horizon is taken as 10 data points (which is 5%), and corresponding one-step ahead predictions are obtained. By using a forecast horizon of 12 data points (implying 5%), a three-step ahead prediction is carried out. The MAE and MSE performance results for all the models in both these cases are presented in Table 1. The original TS is shown in Fig. 6. The predictions for the one-step ahead forecast and three-step ahead forecast are shown in Figs. 7 and 8 respectively. The MA-filter based hybrid ARIMA-ANN model outperformed the others in terms of both MAE and MSE.

Table 1
Performance comparison for TSD1

| Model | One-step-ahead | | Three-step-ahead | |
|---|---|---|---|---|
| | MAE | MSE | MAE | MSE ($\cdot 10^3$) |
| ARIMA | 6.9352 | 70.6029 | 7.1707 | 72.9144 |
| ANN | 6.5810 | 64.8243 | 6.3713 | 61.7111 |
| Zhang | 4.6518 | 44.8343 | 6.1732 | 50.2933 |
| Multiplicative | 4.9226 | 45.2739 | 6.6600 | 63.3805 |
| Khashei and Bijari | 7.7572 | 85.4724 | NA | NA |
| **MA-filter based** | **2.9870** | **13.4466** | **5.3093** | **42.2978** |

### 5.3. Results for TSD2

TSD2 comprises of 100 points, each indicating the number of packets transmitted. The forecast horizon is taken as 10%, which implies 10 data points, and one-step ahead predictions are obtained for these points. A three-step ahead prediction is carried out by using a forecast horizon of 12 which is again nearly 10%. The MAE and MSE performance results for all the models in both these cases are presented in Table 2. The original TS is shown in Fig. 9. The predictions for the one-step ahead forecast and three-step ahead forecast are shown in Figs. 10 and 11 respectively. It is noticed that the MA-filter based hybrid ARIMA-ANN model outperformed the others in terms of both MAE and MSE.

Table 2
Performance comparison for TSD2

| Model | One-step-ahead | | Three-step-ahead | |
|---|---|---|---|---|
| | MAE | MSE | MAE | MSE ($\cdot 10^3$) |
| ARIMA | 14.3987 | 313.6585 | 14.4933 | 332.1901 |
| ANN | 9.2766 | 129.5465 | 11.0087 | 190.3712 |
| Zhang | 9.7403 | 175.3520 | 10.2397 | 169.2432 |
| Multiplicative | 9.4355 | 161.6842 | 13.8069 | 303.3857 |
| Khashei and Bijari | 16.5214 | 386.2702 | NA | NA |
| **MA-filter based** | **6.2872** | **57.6054** | **8.4071** | **102.2310** |

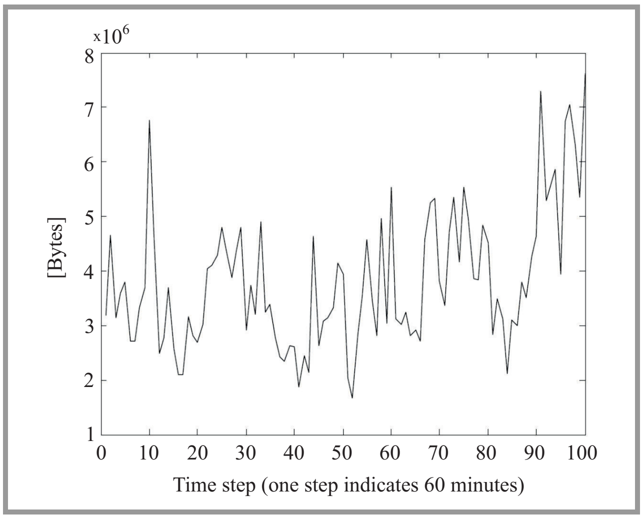


***Fig. 9.*** Actual Internet traffic TSD sampled at 60 min.

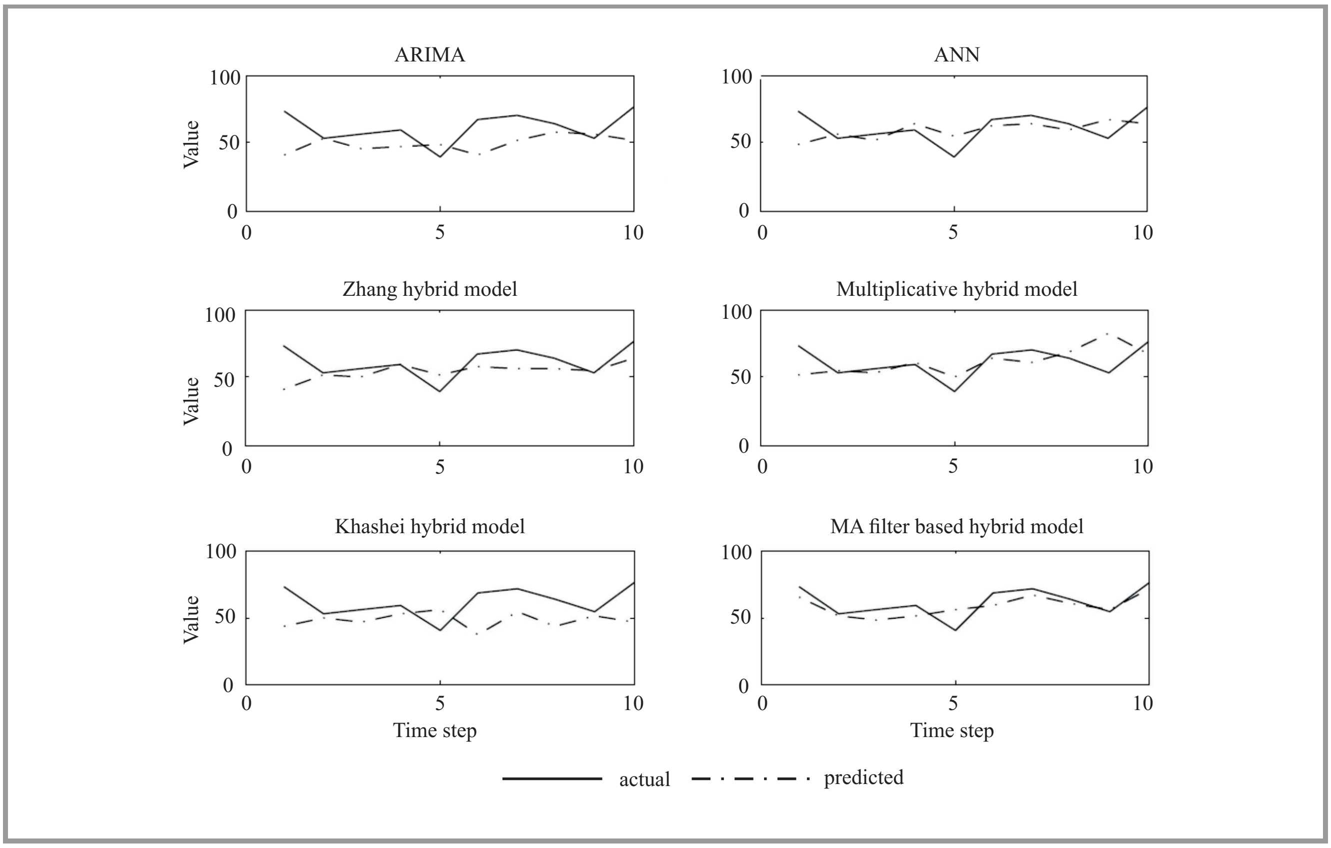


***Fig. 10.*** One-step ahead predictions for TSD2.

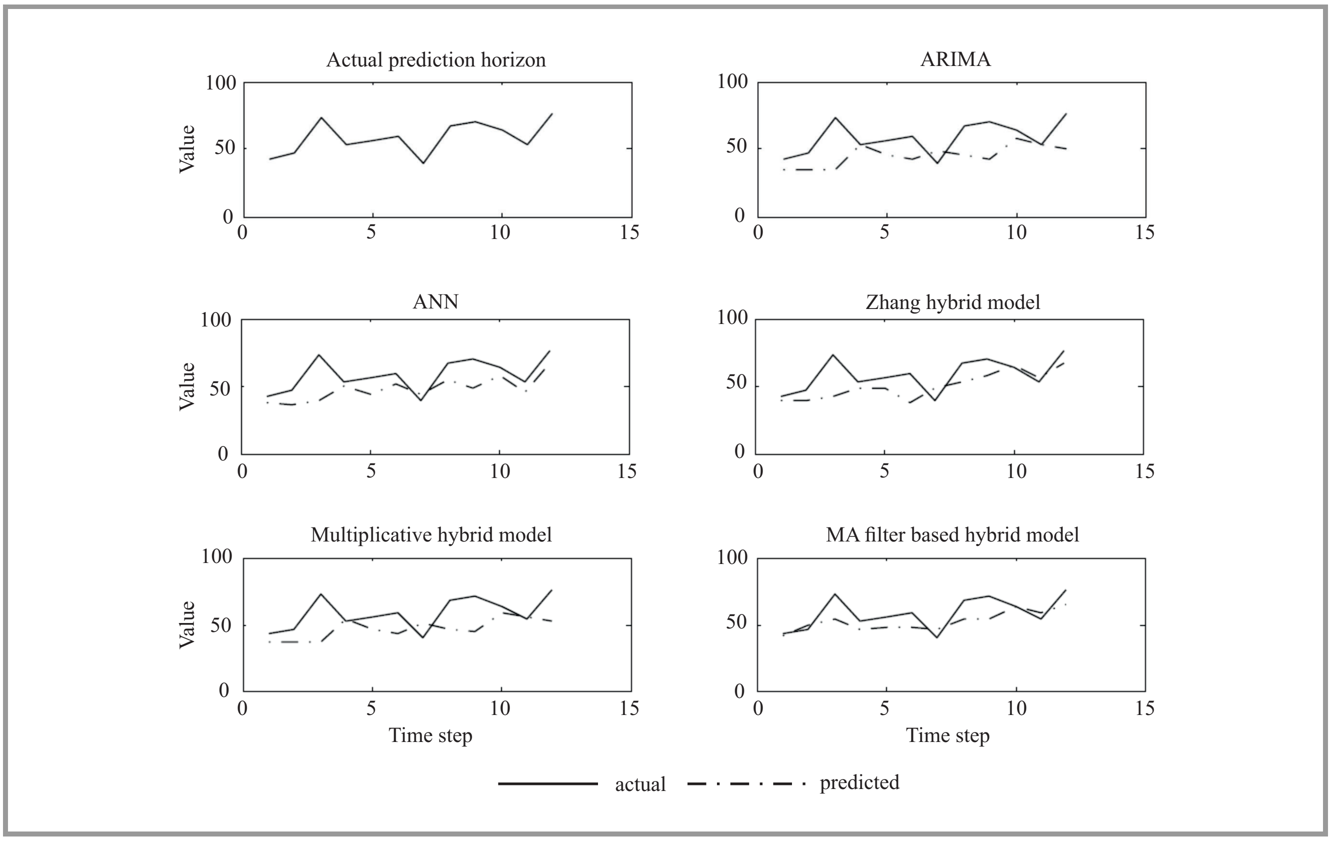


***Fig. 11.*** Three-step ahead predictions for TSD2.

## 6. Conclusion

In this paper, for the prediction of Internet traffic TSD which is highly volatile in nature, the applicability of various prediction models is explored. The models considered in the study are ARIMA, ANN, Zhang's hybrid ARIMA-ANN, Khashei and Bijari's hybrid ARIMA-ANN, multiplicative ARIMA-ANN, MA-filter based hybrid ARIMA-ANN. Both one-step ahead and multi-step ahead predictions are carried out. The error performance measures, MAE and MSE are used to evaluate the model accuracy.

Two traffic TSD series, one with 30 min sampling and 200 data points, other with 60 min sampling and 100 data points are used in the investigation. The prediction results in all the cases showed that the MA filter based hybrid ARIMA-ANN model outperformed all the other models discussed in this paper, in terms of both MAE and MSE and hence is suitable for predicting Internet traffic data more accurately.

## References


[1] J. Contreras, R. Espinola, F. Nogales, and A. Conejo, "ARIMA models to predict next-day electricity prices", *IEEE Trans. on Power Syst.*, vol. 18, no. 3, pp. 1014–1020, 2003.

[2] E. Gonzalez-Romera, M. Jaramillo-Moran, and D. Carmona-Fernandez, "Monthly electric energy demand forecasting based on trend extraction", *IEEE Trans. on Power Syst.*, vol. 21, no. 4, pp. 1946–1953, 2006.

[3] K. Suresh and S. Krishna Priya, "English Forecasting sugarcane yield of tamilnadu using ARIMA models", *English Sugar Tech*, vol. 13, no. 1, pp. 23–26, 2011.

[4] J.-J. Wang, J.-Z. Wang, Z.-G. Zhang, and S.-P. Guo, "Stock index forecasting based on a hybrid model", *Omega*, vol. 40, no. 6, pp. 758–766, 2012.

[5] E. Cadenas and W. Rivera, "Wind speed forecasting in three different regions of Mexico, using a hybrid ARIMA-ANN model", *Renewable Energy*, vol. 35, no. 12, pp. 2732–2738, 2010.

[6] C. Babu and B. Reddy, "Predictive data mining on average global temperature using variants of ARIMA models", in *Proc. 2012 Int. Conf. Adv. Engin. Sci. and Manag. ICAESM 2012*, Tamil Nadu, India, 2012, pp. 256–260.

[7] U. Orhan, "Real-time CHF detection from ECG signals using a novel discretization method", *Comp. in Biology and Medicine*, vol. 43, no. 10, pp. 1556–1562, 2013.

[8] D. Singhal and K. Swarup, "Electricity price forecasting using artificial neural networks", *Int. J. Elec. Power & Energy Syst.*, vol. 33, no. 3, pp. 550–555, 2011.

[9] W.-S. Chen and Y.-K. Du, "Using Neural Networks and data mining techniques for the financial distress prediction model", *Expert Syst. with Appl.*, vol. 36, no. 2, part 2, pp. 4075–4086, 2009.

[10] C. H. F. Toro, S. G. Meire, J. F. Gálvez, and F. Fdez-Riverola, "A hybrid artificial intelligence model for river flow forecasting", *Applied Soft Computing*, vol. 13, no. 8, pp. 3449–3458, 2013.

[11] B. R. Chang and H. F. Tsai, "Novel hybrid approach to data-packet-flow prediction for improving network traffic analysis", *Appl. Soft Comput.*, vol. 9, no. 3, pp. 1177–1183, 2009.

[12] J. Reyes, A. Morales-Esteban, and F. Martínez-Álvarez, "Neural networks to predict earthquakes in Chile", *Appl. Soft Comput.*, vol. 13, no. 2, pp. 1314–1328, 2013.

[13] G. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model", *Neurocomput.*, vol. 50, no. 0, pp. 159–175, 2003.

[14] M. Khashei and M. Bijari, "A novel hybridization of Artificial Neural Networks and ARIMA models for time series forecasting", *Appl. Soft Comput.*, vol. 11, no. 2, pp. 2664–2675, 2011.

[15] H. H. Arash Miranian and M. Abdollahzade, "Day-ahead electricity price analysis and forecasting by singular spectrum analysis", *IET Gener. Transmiss. & Distrib.*, vol. 7, no. 4, pp. 337–346, 2013.

[16] D. Ömer Faruk, "A hybrid neural network and ARIMA model for water quality time series prediction", *Engin. Appl. of Artif. Intell.*, vol. 23, no. 4, pp. 586–594, 2010.

[17] L. Wang, H. Zou, J. Su, L. Li, and S. Chaudhry, "An ARIMA-ANN hybrid model for time series forecasting", *Wiely-Syst. Res. and Behav. Sci.*, vol. 30, pp. 244–259, 2013.

[18] C. N. Babu and B. E. Reddy, "A moving-average-filter-based hybrid arima-ann model for forecasting time series data", *Applied Soft Computing*, vol. 23, pp. 27–38, 2014 [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1568494614002555

[19] Y. Sai, Z. Yuan, and K. Gao, "Mining stock market tendency by RS-based support vector machines", in *Proc. IEEE Int. Conf. Granular Comput. GrC 2007*, Fremont, CA, USA, 2007, pp. 659–664 [Online]. Available: http://dx.doi.org/10.1109/GRC.2007.99

[20] M. R. Hassan, "A combination of Hidden Markov Model and fuzzy model for stock market forecasting", *Neurocomput.*, vol. 72, no. 16–18, pp. 3439–3446, 2009.

[21] G. E. P. Box and G. Jenkins, *Time Series Analysis, Forecasting and Control*. Holden-Day, Inc., 1990.

[22] Q. Yao and P. J. Brockwell, "Gaussian maximum likelihood estimation for ARMA models I: time series", *J. Time Series Anal.*, vol. 27, no. 6, pp. 857–875, 2006.

[23] S. O. Haykin, *Neural Networks and Learning Machines*, 3rd ed. Prentice Hall, 2008.

[24] CAIDA, 2014 [Online]. Available: http://www.caida.org/data/



**C. Narendra Babu** graduated in BE (CSE) from Adichunchanagiri Institute of Technology, Chikmagalur in 2000. He received his M.Sc. in M.Tech. (CSE) from M. S. Ramaiah Institute of Technology, Bangalore in 2004. He submitted his Ph.D. thesis to JNT University, Anantapuramu. He is an Associate Professor in the Department of ISE, Reva Institute of Technology and Management, Bangalore. His research interests include time series data analysis and mining, soft computing. He has published four papers in reputed international journals and two international conferences, also received best author award from IEEE-ICAESM international conference in 2012 held at Nagapattinam. He is a member of IEEE from 2009.

E-mail: narendrababu.c@gmail.com
Department of Computer Science and Engineering
JNT University College of Engineering
Anantapuramu, India

**B. Eswara Reddy** graduated in B.Tech. (CSE) from Sri Krishna Devaraya University in 1995 and obtained M.Sc. degree from JNT University, Hyderabad, in 1999. He received Ph.D. in Computer Science and Engineering from JNT University, Hyderabad, in 2008. He is as Professor of CSE Dept. and Coordinator for Master of Science in Information Technology (MSIT) program, JNTUA in collaboration with Consortium of Institutions of Higher Learning (CIHL). He has written more than 50 publications in various journals and conferences, and two books. He has received University Grants Commission-Major Research Project (UGC-MRP) titled "Cloud computing framework for rural health care in Indian scenario". His research interests include cloud computing, pattern recognition and image analysis, data warehousing and mining, and software engineering. He is a life member of ISTE, IE, ISCA, IAENG and member of CSI, IEEE.

E-mail: eswar.cse@jntua.ac.in
Department of Computer Science and Engineering
JNT University College of Engineering
Anantapuramu, India