CHAPTER 4

HYBRID ARIMA-ANN MODELS FOR INTERNET TRAFFIC TSD

Chapter 4

HYBRID ARIMA-ANN MODELS FOR INTERNET TRAFFIC TSD

This chapter investigates the prediction performance of various hybrid ARIMA-ANN models, when they are applied to internet traffic TSD. The hybrid ARIMA-ANN models compared in this study are our proposed MA filter based hybrid ARIMA-ANN model, Zhang's model, Khashei and Bijari's model, hybrid model of Wang et. al. These hybrid models are applied for both one-step ahead and multi-step ahead prediction cases on internet traffic data sampled at every 30 minutes and 60 minutes. 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.

The chapter is organized as follows. In section 4.1, the characteristics of internet traffic data and corresponding smoothened TSD are discussed. In section 4.2, the first internet TSD sampled every 30 minutes is considered and the prediction performance results are presented. In section 4.3, the second internet TSD sampled every 60 minutes is considered and the prediction performance results are presented. Finally the chapter summary is presented in section 4.4.

4.1 Internet Traffic data

TSD forecasting has its applications in various domains like agricultural, climatic, econometric, financial and communication. With the growing telecommunication sector, the present generation looks forward to use high speed internet services. The service providers should be able to effectively distribute their resources for their continued services. Internet traffic data forecasting helps the service providers to route their 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 apriori 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.

To forecast internet traffic TSD, instead of individual ARIMA (linear) model or ANN (non-linear) model, progressing in the direction of combining the benefits of both ARIMA and ANN models, hybrid ARIMA-ANN models can be more appropriate. For this purpose, in this chapter four different hybrid models are considered for application to internet traffic TSD. These are:

1. The hybrid ARIMA-ANN model proposed by Zhang [5], 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.

2. The hybrid ARIMA-ANN model proposed by Khashei and Bijari [7], which gave better performance than that proposed by Zhang in [5].

3. The multiplicative hybrid ARIMA-ANN model proposed by Li Wang et.al [79], which is in contrast to the additive model of Zhang and whose results showed that it is no less in comparison to the additive Zhang model.

4. The moving average filter based hybrid ARIMA-ANN model proposed and discussed in chapter 1.

The internet traffic TSD predictions for both one-step ahead and multi-step ahead cases are obtained using individual ARIMA, ANN models, and the hybrid ARIMA-ANN models mentioned above. From the prediction performance results obtained for two different internet traffic TSD, the best model was identified as the proposed moving average filter based hybrid ARIMA-ANN model. The raw TSD obtained is sampled for every one second. This data is processed to obtain the number of packets transmitted for every 30 minutes and for every 60 minutes separately and two different TSD are obtained. On each of these TSD the detailed results obtained are given in the consecutive sections. The ANN model is represented as $N^{x,y,z}$, where x is the number of input nodes, y is the size of the hidden layer, and z is the number of output nodes. In this work, $z = 1$ was chosen.

4.2 Results for internet traffic data sampled every 30 minutes

This TSD comprises of 200 data values, each indicating the number of packets transmitted every 30 minutes. Consider one-step ahead forecasting. The forecast horizon is taken as 10 data points, which is 5% of the total data. First ARIMA model orders p, d, q are found using the ACF and PACF plots discussed in Chapter 3. Accordingly the model is ARIMA$(2, 0, 5)$. Using Box-Jenkins approach, the model coefficients are found out, which are given in Table 4.1. The ANN model is fit using the LM training algorithm and the best model was obtained as $N^{5,10,1}$. For Zhang's model, since it is a hybrid of ARIMA and ANN, the ARIMA model order is found as ARIMA$(2, 0, 5)$ and the ANN has $N^{5,14,1}$ architecture. The ARIMA model coefficients are same as those in Table 4.1. For Khashei-Bijari hybrid model, the ARIMA is again of ARIMA$2, 0, 5$, but ANN architecture is $N^{(4+10),14,1}$, where $(4 + 10)$ indicates 4 input delays and 10 feedback delays. For Wang et. al, multiplicative model, the ARIMA model order is ARIMA$(2, 0, 5)$ and ANN has $N^{10,14,1}$ architecture. For the proposed MA-filter based hybrid model, the ARIMA has an order of ARIMA$(10, 2, 0)$ and the ANN architecture used is $N^{5,14,1}$. The model order of ARIMA is different from the other hybrids because it uses decomposition prior to fitting ARIMA model unlike the other models. The

Table 4.1: ARIMA$(2, 0, 5)$ model parameters

Parameter	Value
Constant	5.389
AR1	0.08
AR2	0.627
MA1	0.229
MA2	-0.38
MA3	-0.033
MA4	-0.085
MA5	0.0326
Variance	34.035

Table 4.2: ARIMA$(10, 2, 0)$ model parameters

Parameter	Value
Constant	0.003
AR1	-0.53
AR2	-0.34
AR3	-0.267
AR4	-0.185
AR5	-0.838
AR6	-0.386
AR7	-0.186
AR8	-0.178
AR9	-0.027
AR10	-0.300
Variance	1.8452

ARIMA model coefficients are shown in Table 4.2. The MAE and MSE performance results for all the models in both these cases are tabulated in Table 4.3. The original TSD is shown in Figure 4.1. The predictions for the one-step ahead forecast are shown in Figure 4.2, which show that the proposed model outperformed all the others in terms of prediction accuracy.

By using a forecast horizon of 12 data points implying 5% of the available data, a three-step ahead prediction is carried out. The MAE and MSE performance results for all the models is shown in Table 4.3 and the predictions are shown in Figure 4.3. From the table and the

Table 4.3: Performance comparison for internet traffic TSD sampled every 30 minutes

	One-step-ahead		Three-step-ahead	
	MAE	MSE	MAE	MSE ($*10^3$)
ARIMA	6.9352	70.6029	7.1707	72.9144
ANN	6.5810	64.8243	6.3713	61.7111
Zhang model	4.6518	44.8343	6.1732	50.2933
Multiplicative model	4.9226	45.2739	6.66	63.3805
Khashei and Bijari model	7.7572	85.4724	NA	NA
MA-filter based model	**2.9870**	**13.4466**	**5.3093**	**42.2978**

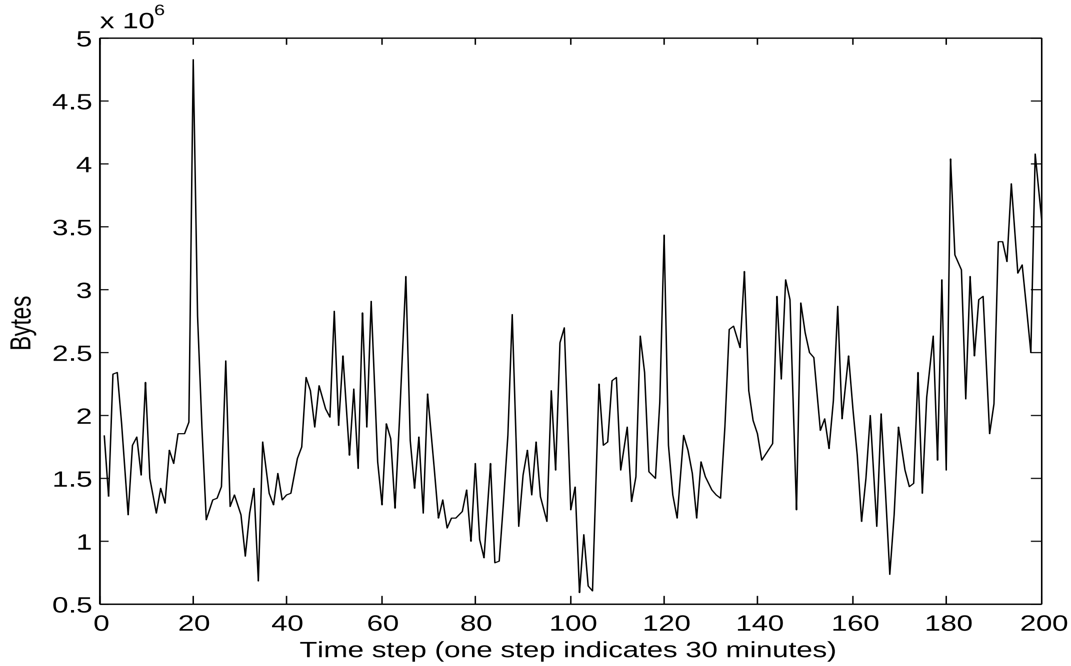

Figure 4.1: Actual internet traffic TSD sampled at 30 minutes

figures shown, it is noticed that the MA-filter based hybrid ARIMA-ANN model outperformed the others in terms of both MAE and MSE.

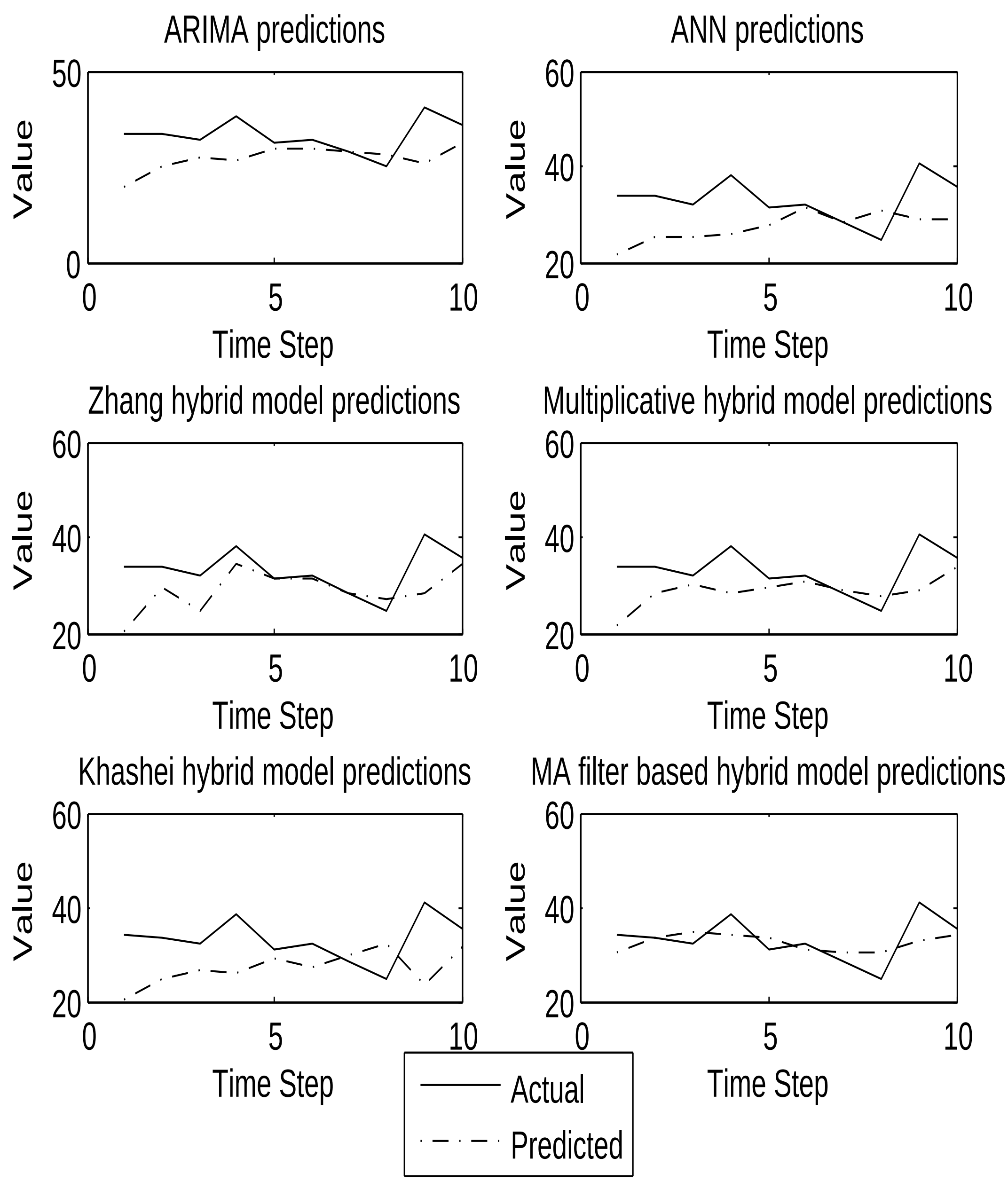

Figure 4.2: One-step ahead predictions for TSD1

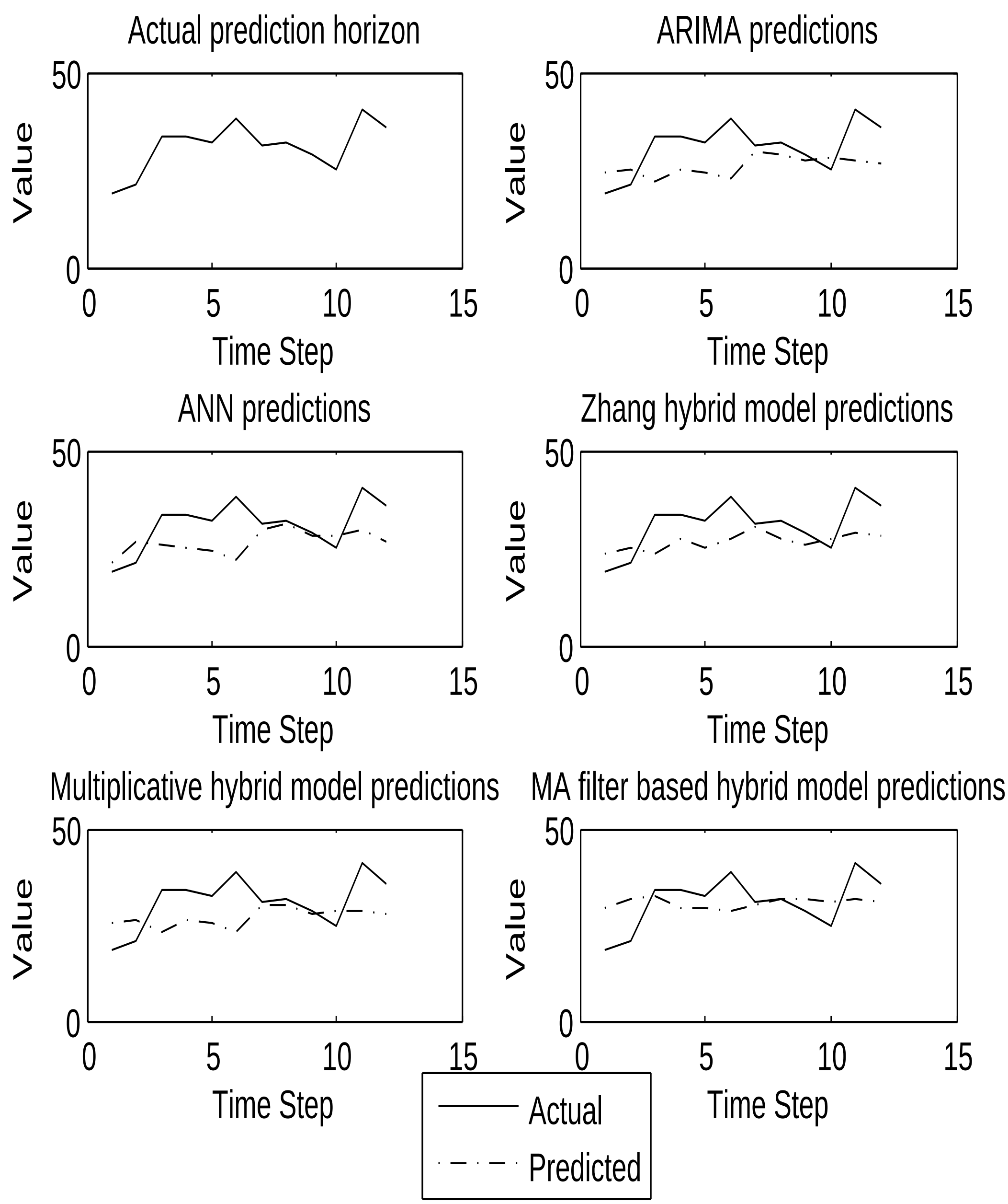

Figure 4.3: Three-step ahead predictions for TSD1

4.3 Results for internet traffic data sampled every 60 minutes

This TSD 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. The ARIMA model has order ARIMA$(1, 0, 1)$. Using Box-Jenkins approach, the model coefficients are found out, which are given in Table 4.4. The ANN model is fit using the LM training algorithm and the best model was obtained as $N^{5,10,1}$. For Zhang's model, the ARIMA model order is found as ARIMA$(1, 0, 1)$ and the ANN has $N^{5,10,1}$ architecture. The ARIMA model coefficients are same as those in Table 4.4. For Khashei-Bijari hybrid model, the ARIMA is again of ARIMA$(1, 0, 1)$, but ANN architecture is $N^{(4+10),10,1}$, where $(4 + 10)$ indicates 4 input delays and 10 feedback delays. For Wang et. al, multiplicative model, the ARIMA model order is ARIMA$(2, 0, 5)$ and ANN has $N^{10,14,1}$ architecture. For the proposed MA-filter based hybrid model, the ARIMA has an order of ARIMA$(8, 1, 0)$ and the ANN architecture used is $N^{7,14,1}$. The model order of ARIMA is different from the other hybrids because it uses decomposition prior to fitting ARIMA model unlike the other models. The ARIMA model coefficients are shown in Table 4.5. A three-step ahead prediction is carried out by using a forecast horizon of 12 which is nearly 10%. The MAE and MSE performance results for all the models in both these cases i.e. for one-step ahead and three-step ahead are tabulated in Table 4.6. The

Table 4.4: ARIMA$(1, 0, 1)$ model parameters

Parameter	Value
Constant	16.25
AR1	0.544
MA1	-0.166
Variance	81.31

Table 4.5: ARIMA$(8, 1, 0)$ model parameters

Parameter	Value
Constant	0.038
AR1	0.323
AR2	-0.085
AR3	0.082
AR4	-0.062
AR5	-0.114
AR6	-0.015
AR7	-0.375
AR8	0.65
Variance	0.632

original TS is shown in Figure 4.4. The predictions for the one-step ahead forecast and three-step ahead forecast are shown in Figure 4.5, and Figure 4.6 respectively. From the table and the figures shown, it is noticed that the MA-filter based hybrid ARIMA-ANN model outperformed all the other models in terms of both MAE and MSE. Hence it is very apt for the prediction of internet traffic TSD than the other hybrids.

Table 4.6: Performance comparison for TSD2

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

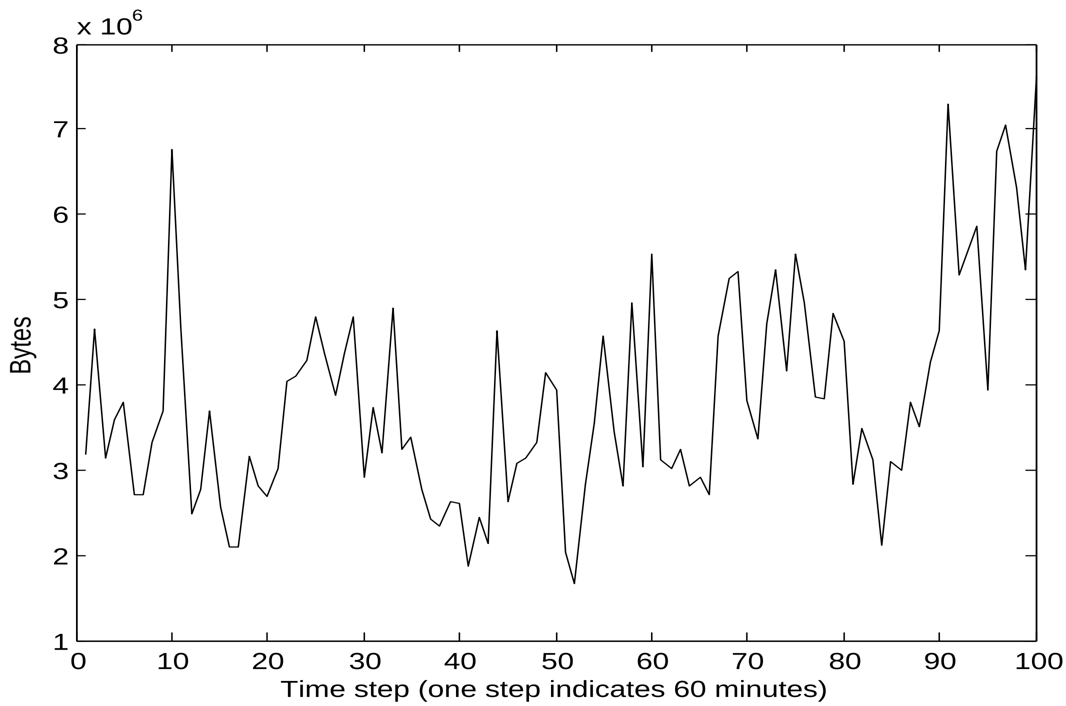

Figure 4.4: Actual internet traffic TSD sampled at 60 minutes

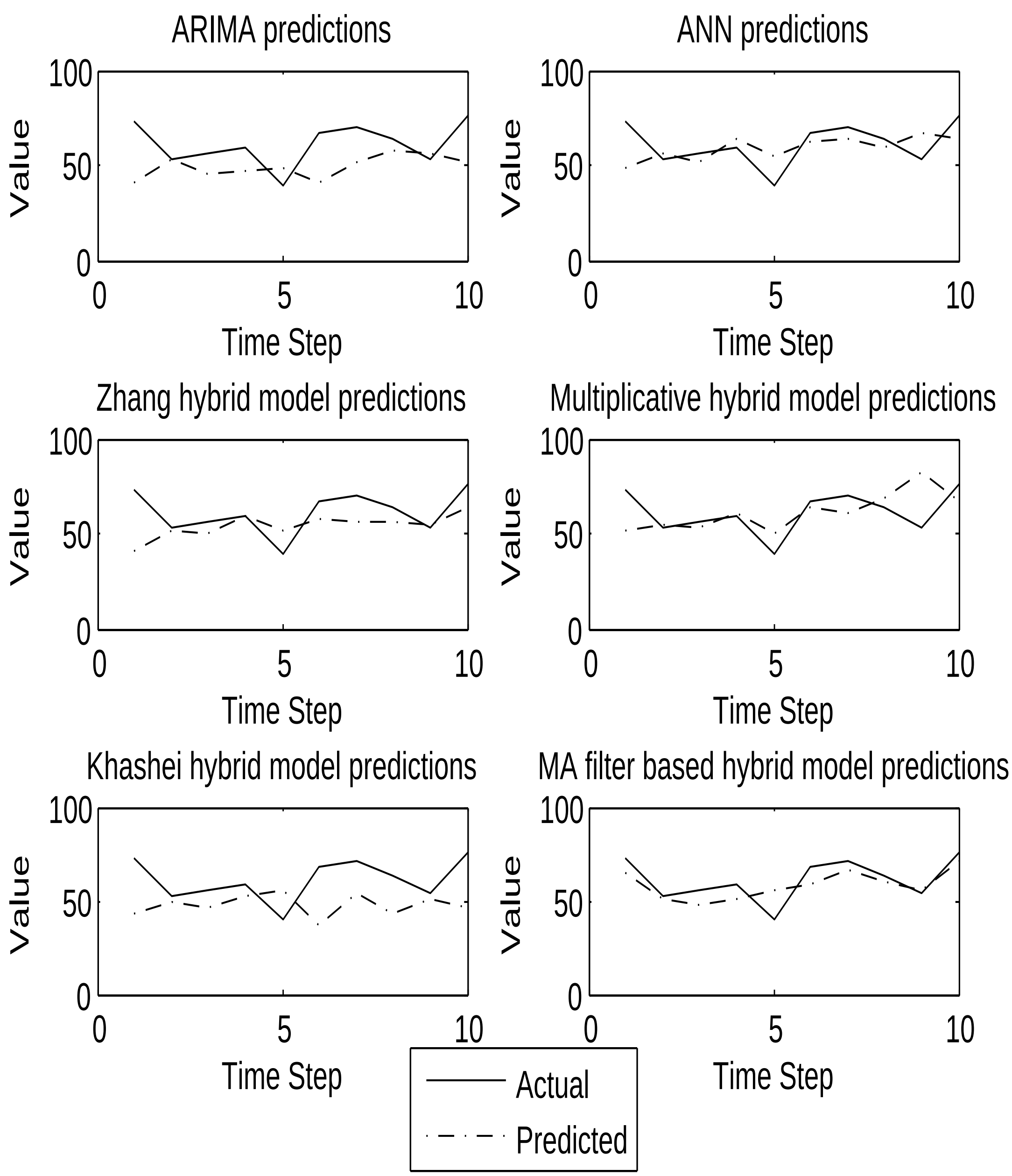

Figure 4.5: One-step ahead predictions for TSD2

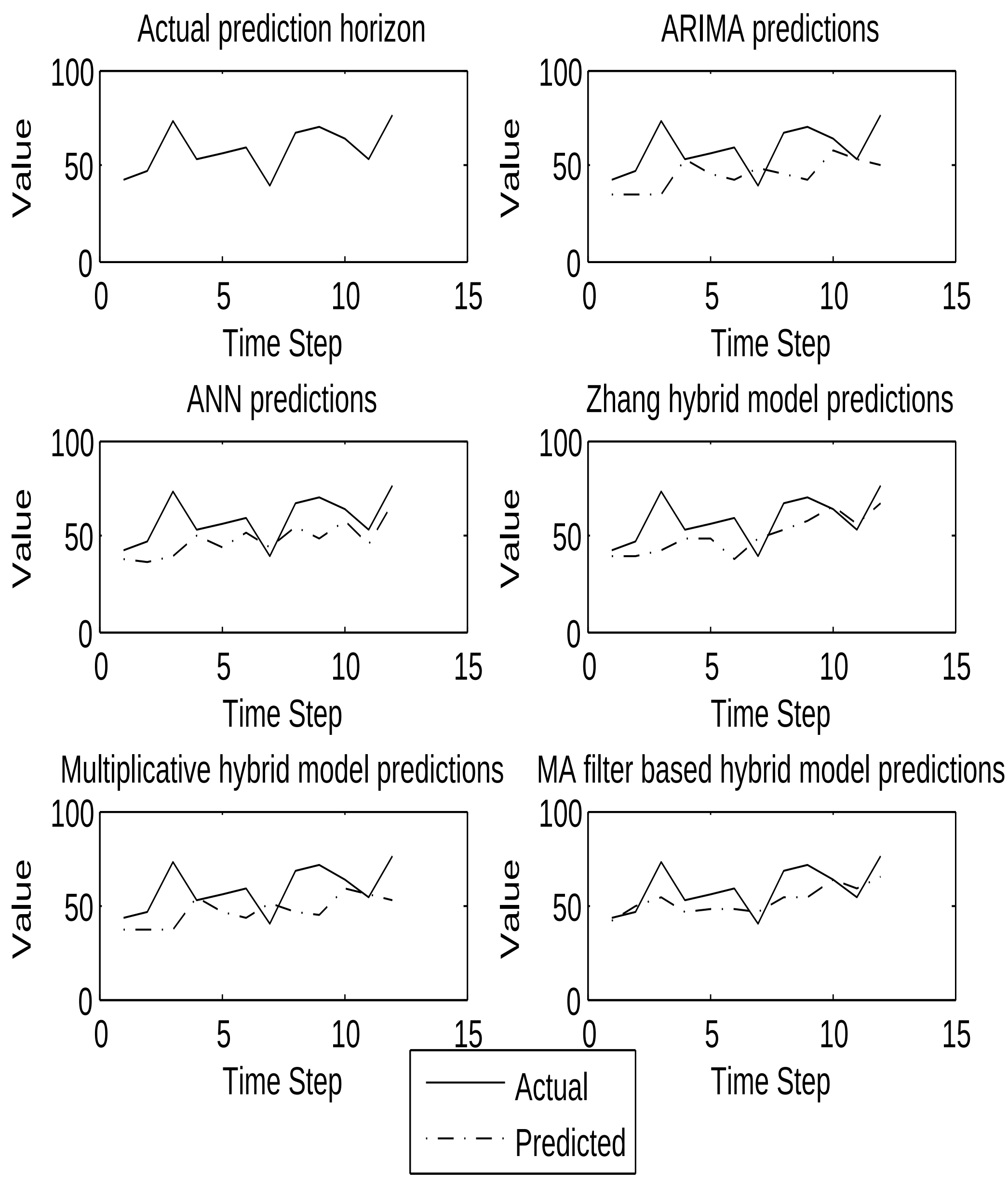

Figure 4.6: Three-step ahead predictions for TSD2

4.4 Summary

In this chapter, 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 proposed in chapter 4. 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 minute sampling and 200 data points, other with 60 minute 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 research work, in terms of both MAE and MSE and hence is suitable for predicting internet traffic data more accurately than the others.