CHAPTER 3

MA-FILTER BASED HYBRID ARIMA-ANN MODEL

Chapter 3

MA-FILTER BASED HYBRID ARIMA-ANN MODEL

This chapter presents a non-linear hybrid ARIMA-ANN prediction model based on a moving-average filter, which is proposed to obtain one step-ahead and multi-step ahead forecasts on TSD. Further, the statistical approach to the proposed model is presented which justifies the performance improvement of the proposed method. This model has been applied on a simulated TSD, and from the performance comparisons with ARIMA, ANN, Zhang's hybrid ARIMA-ANN, Khashei and Bijari's hybrid ARIMA-ANN, it is understood that the proposed model gives better prediction accuracy than these models. To further confirm this performance improvement, the proposed model is applied to obtain both one-step ahead and multi-step ahead forecasts on sunspot TSD, financial TSD and electricity price TSD. It is observed that in all these cases, the proposed method gave more accurate forecasts than the other hybrids [5], Khashei and Bijari [7].

The chapter is organized as follows. Section 3.1 presents the brief overview on some of the existing hybrid ARIMA-ANN models. In section 3.2, motivation and statistical approach to the proposed model, its mathematical details with corresponding algorithm steps and flow chart, and finally advantages of the proposed model are all illustrated.

In section 3.3, the detailed application of the proposed model along with other hybrids and individual models is presented on a simulated TSD and experimental TSD, namely sunspot TSD, electricity price TSD and financial TSD. Finally the chapter summary is presented in section 4.4.

3.1 Overview of existing hybrid ARIMA-ANN models

Often, a given TSD may have both linear and nonlinear characteristics. So, a suitable combination of both linear and nonlinear models, yields a more accurate prediction model than individual models for forecasting TSD of different origin. So, hybrid models using both ANN and ARIMA methods are better than individual ARIMA or ANN models for obtaining accurate predictions. Many hybrid ARIMA-ANN models which exist in the literature incorporate the following strategy: Given a TSD, ARIMA model is directly fit on the data. The error between the given TSD and the predicted TSD is considered as a nonlinear component, and this error TSD is modeled using ANN in different ways. Some such hybrid models considered in this thesis are those of Zhang [5] and Khashei and Bijari [7], which are illustrated below.

3.1.1 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 (3.1).

$$y_t = L_t + N_t \tag{3.1}$$

According to the modeling procedure of Zhang, on the given TSD y_t, first, ARIMA model is fit and the linear predictions are obtained, which are notated as $\hat{L}_t$, and are given in (3.2). In (3.2), the ARIMA model coefficients are $a_1, a_2, \ldots, a_p, b_1, b_2, \ldots, b_q$, which implies the fit ARIMA model has AR model order p and MA model order q. The methodology of ARIMA model is detailed in chapter 1.

$$\hat{L}_t = a_1 y_{t-1} + ... + a_p y_{t-p} + b_1 e_{t-1} + ... + b_q e_{t-q} + e_t \tag{3.2}$$

After obtaining the predictions from ARIMA, they are subtracted from the given TSD, and the difference series is obtained as in (3.3). According to Zhang, this difference series or error series comprises of non-linear variations because ARIMA model can fit linear variations accurately.

$$n_t = y_t - \hat{L}_t \tag{3.3}$$

On this error series notated as n_t, given in (3.3), ANN model is fit and the predictions $\hat{N}_t$ are obtained using (3.4). In (3.4), $\hat{N}_t$ represents the predicted non-linear error series and f is a non-linear function of previous error values. v_t is the white noise component used in the ANN modeling.

The modeling steps of ANN are detailed in Chapter I.

$$\hat{N}_t = f(n_t, n_{t-1}, ..., n_{t-A}) + v_t \tag{3.4}$$

The final predictions of Zhang's hybrid ARIMA-ANN model are obtained by summing the ARIMA predictions in (3.2) and ANN predictions in (3.4), which is given in (3.5). This model is suitable for both one-step ahead and multi-step ahead predictions. It is shown to be better than individual models in terms of prediction accuracy. This hybrid model is sketched in Figure 3.1.

$$\hat{y}_t = \hat{L}_t + \hat{N}_t \tag{3.5}$$

In the work of Zhang, [5], this hybrid model along with individual ARIMA and ANN models are applied on sunspots data, Canadian lynx data and financial TSD. In all the cases for one-step ahead prediction, it is shown that the hybrid model of Zhang gave more accurate predictions than the individual ARIMA or ANN models.

3.1.2 Khashei and Bijari's hybrid ARIMA-ANN model

In 2010, Khashei and Bijari proposed a new hybrid ARIMA-ANN model for forecasting TSD. Similar to Zhang's model, their model also assumes that any TSD has linear and non-linear components (3.1). But the methodology adopted in prediction is different. According to this model, on the given TSD, an ARIMA model is fit to obtain one linear forecast on the TSD using (3.2). Then the past values of given TSD, present linear forecast obtained from ARIMA, and past error data are all given as input to the ANN. The ANN gets trained and once the model is validated, the final one-step forecast of the given TSD is directly obtained (3.6). Note that unlike Zhang model, here, ANN gives f, which is a non-linear function of $\hat{L}_t$, previous linear component TSD values, L_{t-A} and previous values of error series n_t given in (3.2). Also, in (3.6), A and B represent the integers determined in the ANN modeling process. It is interesting to note that using this model, directly the final predictions are obtained, without necessitating the summation of linear and non-linear predictions. However the model complexity is comparatively higher than that of the Zhang's hybrid model.

$$\hat{y}_t = f(\hat{L}_t, L_{t-1}, L_{t-2}, \ldots, L_{t-A}, n_{t-1}, \ldots n_{t-B}) + v_t \tag{3.6}$$

In the work of Khashei and Bijari, [7] this hybrid model is applied on sunspots data, lynx data and some financial TSD. Accordingly, this model accuracy is found to be better than that given by hybrid model of Zhang, individual ARIMA or ANN models, in the case of one-step ahead prediction. However its model accuracy is no better than that of Zhang for multi-step forecasting, because in that case, the past predictions are used as inputs instead of past original values, and hence the model accuracy degrades. The model is illustrated in Figure 3.2.

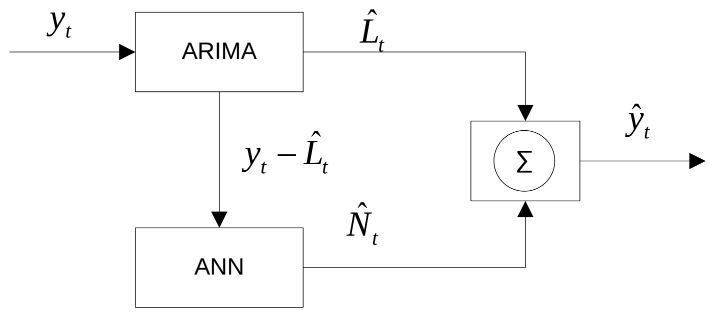

Figure 3.1: Zhang's hybrid ARIMA-ANN model

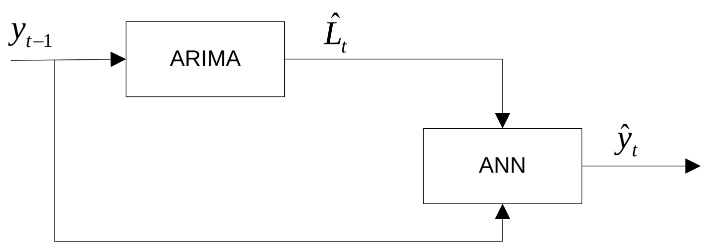

Figure 3.2: Khashei and Bijari's hybrid ARIMA-ANN model

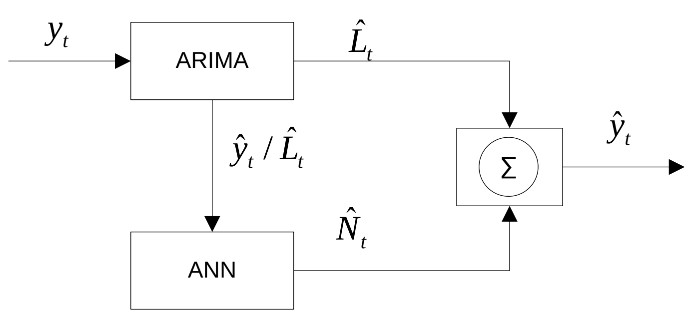

Figure 3.3: Multiplicative hybrid ARIMA-ANN model

3.1.3 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 shown in (3.7) unlike the additive nature assumed by Zhang in (3.1). In (3.7), L_t is the linear and N_t is the non-linear component.

$$y_t = L_t N_t \tag{3.7}$$

The given TSD y_t is modeled using ARIMA as in (3.2), similar to that in Zhang model. Dividing the original TSD by the predictions $\hat{L}_t$, we obtain the non-linear TSD as given by (3.8). It is interesting to note that the substraction in (3.3) is replaced by division in this model.

$$n_t = \frac{y_t}{\hat{L}_t} \tag{3.8}$$

The series n_t is modeled and predicted using ANN. The obtained non-linear predictions $\hat{N}_t$ in (3.3) and linear predictions $\hat{L}_t$ are multiplied to obtain the final model forecasts as illustrated in (3.9). The block diagram of this model is as shown in Figure 3.3.

$$\hat{y}_t = \hat{L}_t \hat{N}_t \tag{3.9}$$

This model is different from Zhang's model, but the obtained predictions have similar accuracy to that of Zhang's model as shown in the work of

Wang et. al. [79]. In [79], using a variety of TSD including sunspot data, lynx data and various other financial data, it is shown that the model has almost similar accuracy as that of Zhang's additive hybrid ARIMA-ANN model.

3.2 Proposed hybrid ARIMA-ANN model

Motivation: Though the existing hybrid models give more accurate predictions than the individual ARIMA and ANN models, there is a scope for further improvement in prediction accuracy if the nature of given TSD is taken into account before application of these models. Hence, in our proposed work, the volatility nature of TSD is explored using moving-average filter, and then on each of the obtained decompositions, a suitable model, ARIMA or ANN is applied. This proposed moving-average filter based hybrid ARIMA-ANN model is applied on a simulated data set and experimental data sets such as sunspot data, electricity price data, and stock market data, and it is observed that for both one-step-ahead and multistep-ahead forecasts, the proposed hybrid model has higher prediction accuracy than the above discussed existing hybrid models.

3.2.1 Statistical approach to the proposed model

The proposed hybrid ARIMA-ANN model is outlined in this section. The technique first decomposes the given data based on the nature of volatility of TSD. Then ARIMA and ANN models are suitably applied. Before describing this technique, we first discuss some interesting facts about

ARIMA sequences, which are used in understanding and characterizing the given data.

In the hybrid methods proposed by Zhang [5] and by Khashei and Bijari [7], the data are assumed to be the sum of linear and nonlinear components. But the given data are not decomposed into linear and nonlinear components; instead, a linear ARIMA model is fit directly to the data and the error sequence thus obtained is assumed to be the nonlinear component. Thus both of these hybrid methods explore and use the fact that the ARIMA model is linear. Ideal ARIMA sequences have many interesting properties, two of which are linearity and stationarity. Some other statistical facts about ARIMA and ARMA sequences are the following: first, the error sequence n_t in ARIMA is Gaussian or normally distributed and white in nature [1], and second, a Gaussian time series represented as a random vector $[y_t\ y_{t-1}\ y_{t-2}\ \ldots]$ is joint-Gaussian in nature [100].

The second statistical fact can be explored further as follows. A stationary Gaussian time series is always stationary in the strict sense [100]. So, assuming that a given ARMA time series is strictly stationary, one possibility is that this series is a Gaussian time series. Usually, after making the given time series data stationary, estimation of the ARIMA model coefficients is performed using GMLE [1]. In this estimation, the model coefficients are obtained as if the given time series were Gaussian. So, if the given stationary time series is truly Gaussian, then the estimated ARIMA model is a better fit. So, it can be concluded that if the time series is stationary in the strict sense, an ARIMA model is more suitable for Gaussian time series data. Then the random vector $[y_t \; y_{t-1} \; y_{t-2} \; \ldots]$ is joint-Gaussian and each random variable y_t is Gaussian distributed.

In general, to diagnose whether a given sequence is normally distributed or not, the Jarque-Bera normality test can be performed. A part of this test checks whether the kurtosis of the sequence, given in (3.10), is 3 or not:

$$\text{kurtosis} = \frac{E\left\{(y - E\{y\})^4\right\}}{\left(E\left\{(y - E\{y\})^2\right\}\right)^2} \tag{3.10}$$

In (3.10), y is the random variable for which the kurtosis is being computed, and E stands for the expectation operation. If the kurtosis value is 3, then the sequence is Gaussian; such sequences were considered as low-volatility data in this research. Sequences that did not have a kurtosis value of 3 were considered as highly volatile data. A highly volatile time series is either leptokurtic or platykurtic in nature, which means that the distribution is non-Gaussian. Thus we can conclude that ARIMA models are suitable for any time series data when the data have a kurtosis value of approximately 3. With this understanding, the proposed model is described below.

Mathematically, the proposed model can be described as follows. The time series data y_t are considered as a sum of a low-volatility component l_t and a high-volatility component h_t, as given in (3.11). After making sure that l_t is stationary, it is modeled as a linear function of past values of the sequence $l_{t-1}, l_{t-2}, \ldots, l_{t-p}$ and the random error sequence $n_t, n_{t-1}, \ldots, n_{t-q}$ using an ARIMA model. This is shown in (3.12), where f is a linear function. Similarly, h_t is expressed as a nonlinear function of $h_{t-1}, h_{t-2}, \ldots, h_{t-N}$ as shown in (3.13), and is modeled using an ANN. In (3.13), g represents the nonlinear function, and ε_t represents the model error. Using the ARIMA-predicted low-volatility component $\hat{l}_t$ and the ANN-predicted high-volatility component $\hat{h}_t$, the predicted time series value $\hat{y}_t$ is obtained as represented in (3.14).

$$y_t = l_t + h_t \tag{3.11}$$

$$\hat{l}_t = f(l_{t-1}, l_{t-2}, \ldots, l_{t-p}, n_t, n_{t-1}, \ldots, n_{t-q}) \tag{3.12}$$

$$\hat{h}_t = g(h_{t-1}, h_{t-2}, \ldots, h_{t-N}) + \varepsilon_t \tag{3.13}$$

$$\hat{y}_t = \hat{l}_t + \hat{h}_t \tag{3.14}$$

3.2.2 Algorithmic steps and flow chart:

The steps of the algorithm for the proposed hybrid model are given below and are represented as a flow chart in Figure 3.4.

1. Using an MA filter, given in (3.15), the given time series data are separated or decomposed into two components such that one of

the components is less volatile and the other is highly volatile. The length of the MA filter, m, is adjusted so that this decomposition is properly achieved. The first decomposition is y_{tr}, given in (3.15), which is the smoothed trend component, and has low volatility. The second decomposition obtained from the MA filter is the residual component, given in (3.16), which has high volatility.

2. The low-volatility component with $k = 3$ is modeled using an ARIMA model and the predictions are obtained as in (3.12).

3. The high-volatility component with $k \mathrel{!}= 3$ is modeled using an ANN and the predictions are obtained as in (3.13).

4. The predictions obtained from steps 2 and 3 are added to obtain the final predictions as in (3.14):

$$y_{tr} = \frac{1}{m} \sum_{i=t-m+1}^{t} y_i \tag{3.15}$$

$$y_{res} = y_t - y_{tr} \tag{3.16}$$

3.2.3 Advantages:

The proposed algorithm can give better accuracy compared with the other hybrid models discussed in the previous section, which first directly fit an ARIMA model to the given data. This can be understood from the following reasoning. A linear sequence cannot be accurately modeled by a nonlinear model, and vice versa. A low-volatility series is

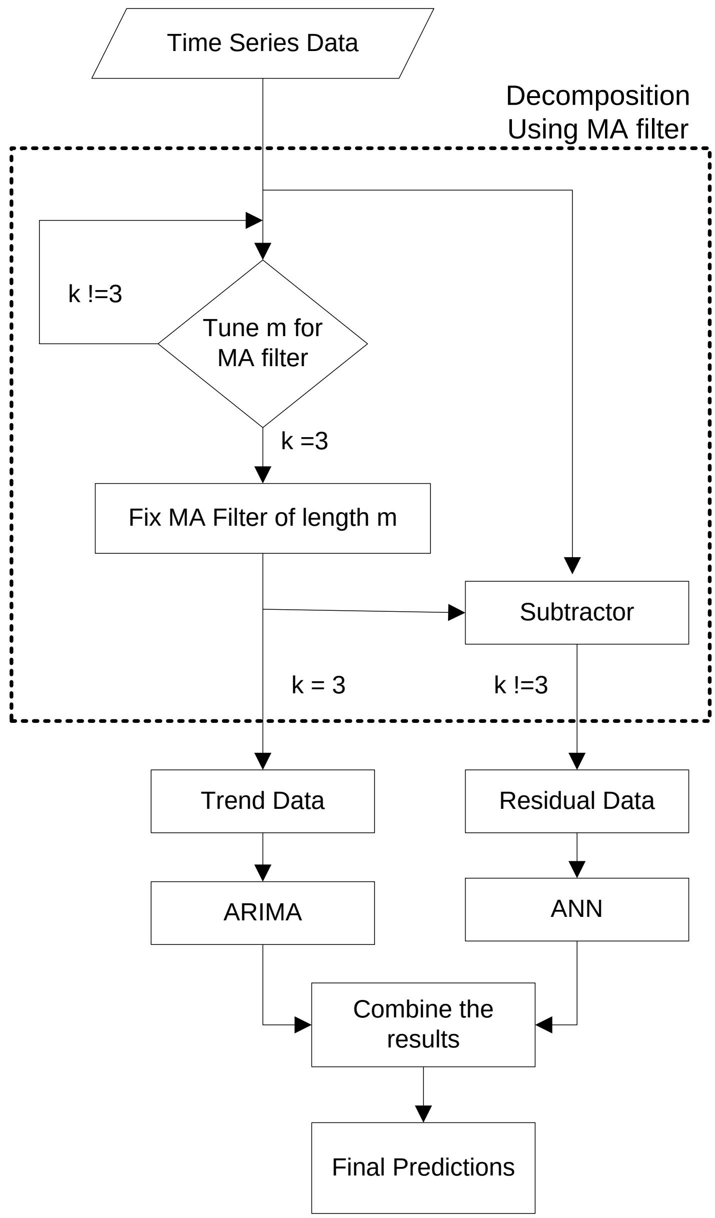

Figure 3.4: MA filter based hybrid ARIMA-ANN model

Gaussian in nature and an ARIMA model suits it better, which implies that it can be modeled accurately using a linear model. Therefore, when the time series is less volatile, it can be considered as a linear sequence. Similarly, if it is highly volatile in nature, it can be considered as a nonlinear sequence. If a linear sequence is modeled by a linear model, the model error will be small. The case for a nonlinear sequence is similar. So, when a given data set is decomposed into low and high-volatility components, which are almost linear and nonlinear components, respectively, the total model error will be small. On the other hand, if an ARIMA model is directly fit to the data, the separation of linear and nonlinear components is not performed. So, there is a chance that part of the linear component will be modeled by a nonlinear model, resulting in an increase in the model error. Hence, the proposed model can give more accurate results than Zhang's and Khashei-Bijari's models. This fact has been verified using simulated and experimental data sets.

3.3 Results and discussion

The proposed method, along with other four modeling techniques considered are applied to simulated data and also to time series data of various kinds. A detailed description of the results is presented in this section. Before any further discussion of the results, however, the performance measures used for comparison of prediction accuracy will be discussed. The two performance measures considered for accuracy comparison are MAE and MSE.

3.3.1 Results for simulated data

A known data set was generated by adding a linear data process AR($2, 0, 0$) to a nonlinear data set simulated by an ANN. The nonlinear 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. The simulated ANN data corresponded to $N^{2,2,1}$. The resultant data were modeled using ARIMA and ANN models, Zhang's hybrid model [5], Khashei and Bijari's hybrid model [7], and the proposed hybrid model.

For the ARIMA modeling, a suitable model order was found using the R software package, and then this model was fit to the data using MATLAB. For the ANN and hybrid modeling, MATLAB was used. The total number of data points taken was 100. In the case of one-step prediction, the forecast horizon was 10. The multistep-ahead prediction performed on the data was a three-step-ahead prediction. For this, the forecast horizon considered was 30. The results for the performance measures obtained with this data are shown in Table 3.1. The actual time series data are shown in Figure 3.5, the predictions for one-step-ahead forecasting are shown in Figure 3.6, and the predictions for three-step-ahead forecasting are shown in Figure 3.7. From Table 3.1 and the plots in Figure 3.6, it can be seen that the proposed method gives better performance than all of the other models used for comparison.

In the case of multistep-ahead prediction, the results from Khashei and Bijari's model are not given, because they were almost same as those

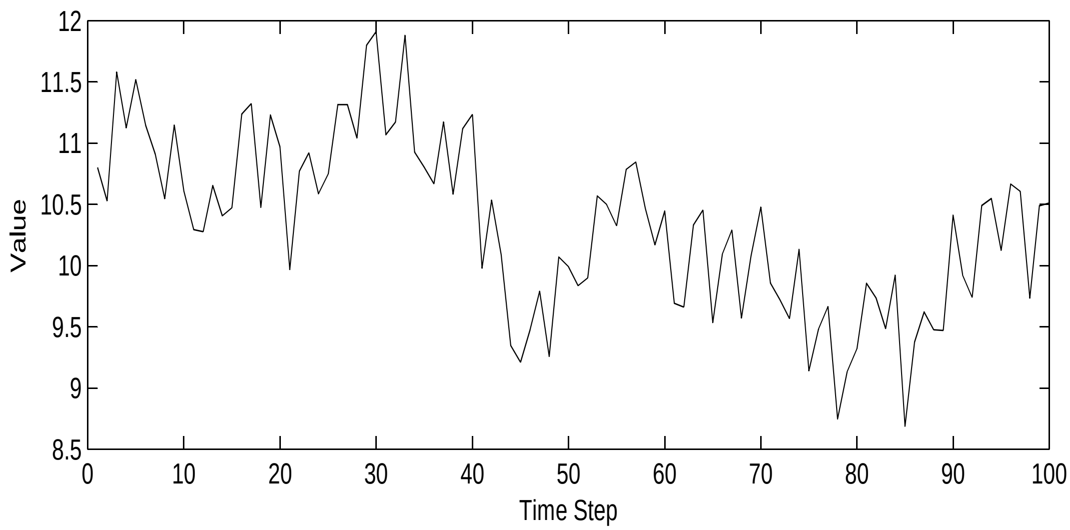

Figure 3.5: Simulated time series data

from Zhang's model. This is because in Khashei and Bijari's model [7], the ANN inputs should be the ARIMA-predicted present value, the past errors, and the past actual data values, for one-step-ahead prediction. If this model has to be used for multistep-ahead prediction, multiple future values have to be predicted. Consider a data set having 100 points. For a five-step-ahead prediction, the 101^{st} to 105^{th} values have to be predicted based on only the first 100 data points. In this case, if Khashei and Bijari's model is used, the 101^{st} point can be predicted. To predict the 102^{nd} point, the ARIMA-predicted 102^{nd} point and the past errors are available, but among the past actual values needed by the model, the 101^{st} actual value will not be known, so the model is not suitable for multistep-ahead prediction. One way to overcome this problem would be to use the 101^{st} model prediction instead of the 101^{st} actual value, but this reduced the model accuracy and it was observed that the model

Table 3.1: Performance comparison for simulated data

	One-step-ahead		Three-step-ahead	
	MAE	MSE	MAE	MSE
ARIMA	0.3925	0.1896	0.4474	0.3070
ANN	0.3582	0.1764	0.3857	0.2372
Zhang model	0.2595	0.0841	0.3487	0.2032
Khashei and Bijari model	0.2365	0.0701	NA	NA
Proposed model	**0.1884**	**0.0507**	**0.2951**	**0.1445**

accuracy was no better than that of Zhang's model. So, for multistep-ahead prediction, the results from this model have not been included, as they do not provide an apt comparison.

From the results in Table 3.1, it can be verified that the proposed model gives better performance than the other models for the simulated data, i.e., a known time series data set. With this success in mind, the discussion now progresses towards applying the model to real time series data sets obtained from various applications. The time series data considered in this research work are sunspot data, electricity price data from the Australian National Electricity Market, and the close prices of stock from the National Stock Exchange, India.

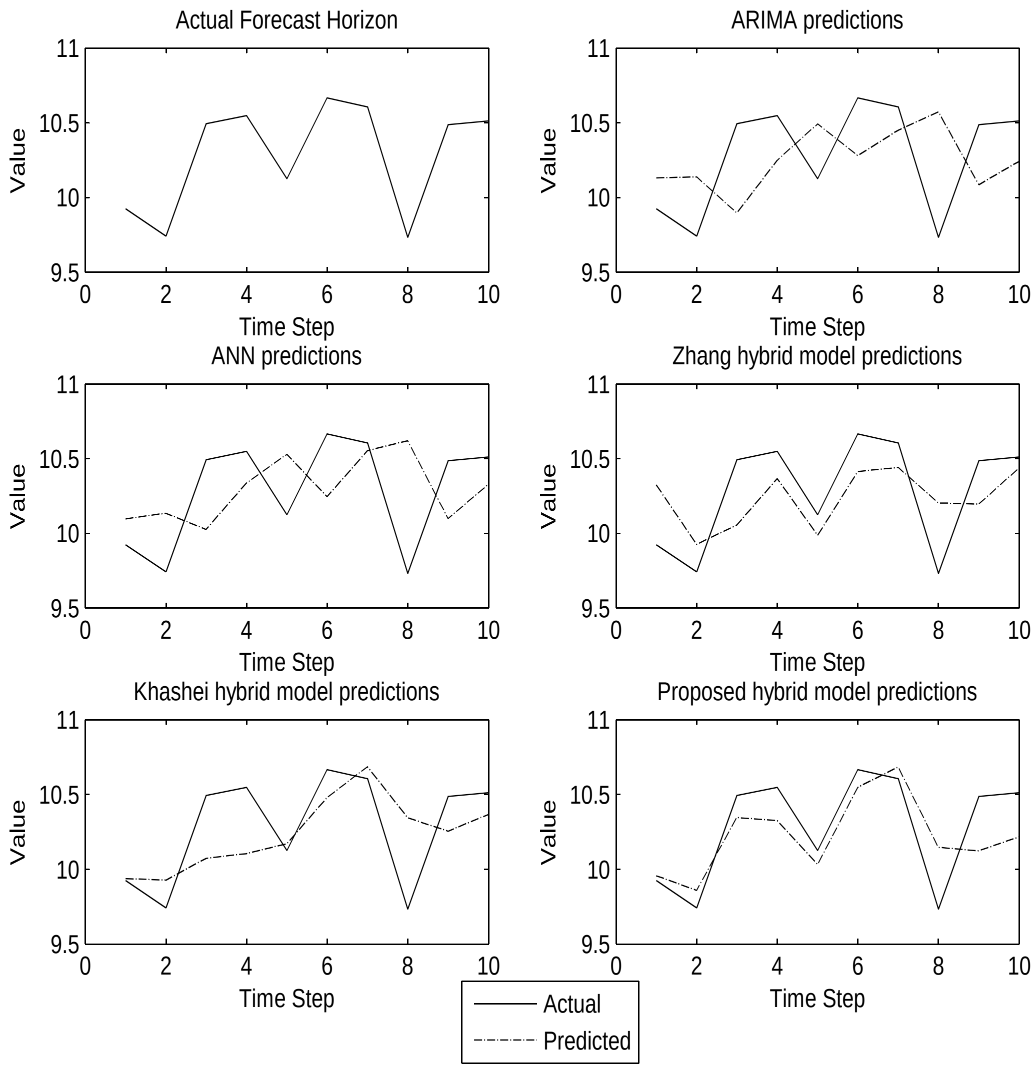

Figure 3.6: One-step-ahead predictions for simulated time series data

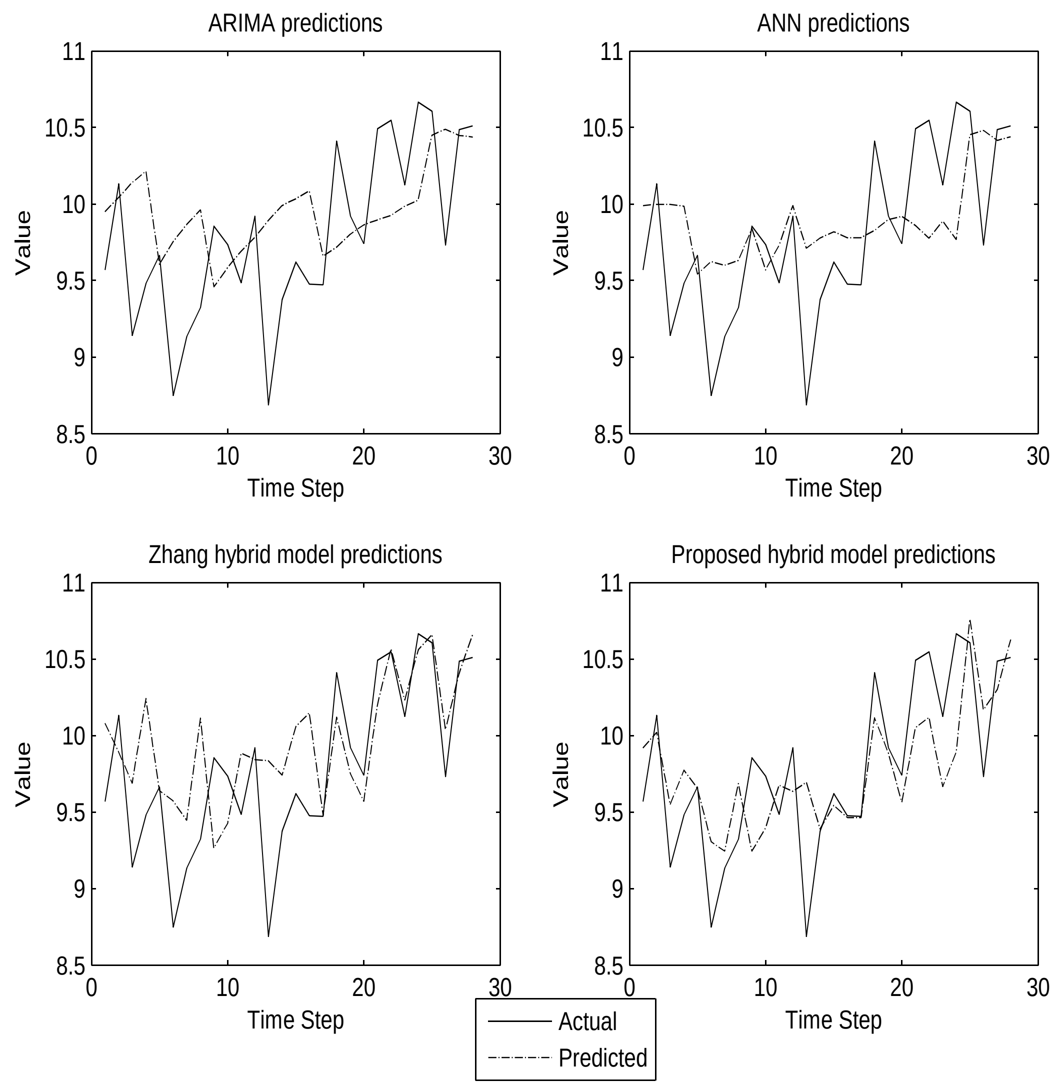

Figure 3.7: Three-step-ahead predictions for simulated data using various models

Table 3.2: ARIMA$(9,0,0)$ model parameters

Parameter	Value
Constant	5.146
AR1	1.15
AR2	-0.384
AR3	-0.143
AR4	0.119
AR5	-0.084
AR6	0.032
AR7	-0.031
AR8	0.056
AR9	0.183
Variance	228.6

3.3.2 Results for sunspot data

Sunspot time series data from 1700 to 1987 were considered for this study; these data were a set of 288 points. For one-step-ahead prediction, the forecast horizon was chosen as 25 data points. The multi-step prediction performed in this case was five-step ahead prediction. The forecast horizon considered was 50. The ARIMA model fit in ARIMA, Zhang's hybrid, Khashei-Bijari hybrid model is ARIMA$(9,0,0)$, whose model parameters are shown in Table 3.2. The ARIMA model fit in the proposed hybrid model is ARIMA$(10,0,0)$, whose model coefficients are shown in Table 3.3. The prediction performance results for all the models are tabulated in Table 3.4 for both of these cases. The original time series data are shown in Figure 3.8. The predictions for the one-step-ahead forecast are shown in Figure 3.9, and those for the five-step-ahead forecast are shown in Figure 3.10. From the table and the figures shown, it can be seen that the proposed method outperforms all of the other models used for comparison in terms of MSE and MAE.

Table 3.3: ARIMA$(10, 0, 0)$ model parameters

Parameter	Value
Constant	0.425
AR1	2.098
AR2	-1.434
AR3	0.14
AR4	0.3469
AR5	-0.287
AR6	0.286
AR7	-0.443
AR8	0.493
AR9	-0.062
AR10	-0.149
Variance	0.469

Table 3.4: Performance comparison for sunspot data

	One-step-ahead		Five-step-ahead	
	MAE	MSE	MAE	MSE ($*10^3$)
ARIMA	14.6630	308.1491	22.0840	1.0993
ANN	14.3513	285.1216	21.7042	0.9384
Zhang model	14.2233	298.5670	20.8829	1.0309
Khashei and Bijari model	13.4053	269.5369	NA	NA
Proposed model	**9.8718**	**155.5646**	**17.7869**	**0.5769**

In the proposed method, when the MA filter was used, the length was fixed at 37. The given time series data had a kurtosis value of 3.6, indicating that it was highly volatile. After using the filter, the smoothed component, which we call the trend component, had a kurtosis of 3, which indicated that it had low volatility and the ARIMA method was suitable for modeling. The residual component obtained from the filter had a kurtosis of 3.2, indicating that it was a highly volatile component and the ANN method was suitable. Thus the proposed model was applied as per the discussion in Section 3.2. Multistep forecasting generally has less accuracy than one-step-ahead forecasting which can be observed from the results tabulated in Table 3.4.

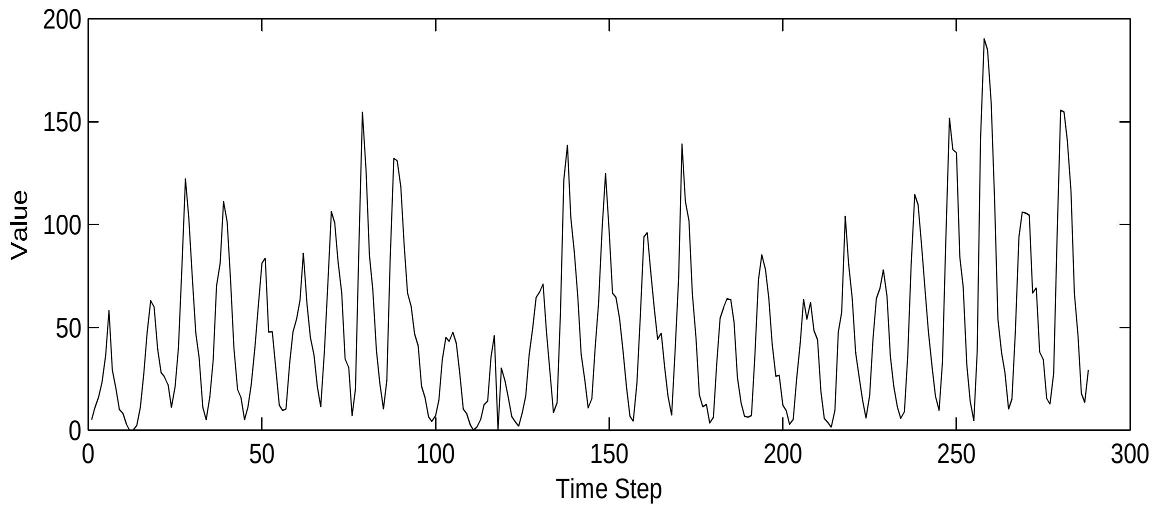

Figure 3.8: Sunspot time series data

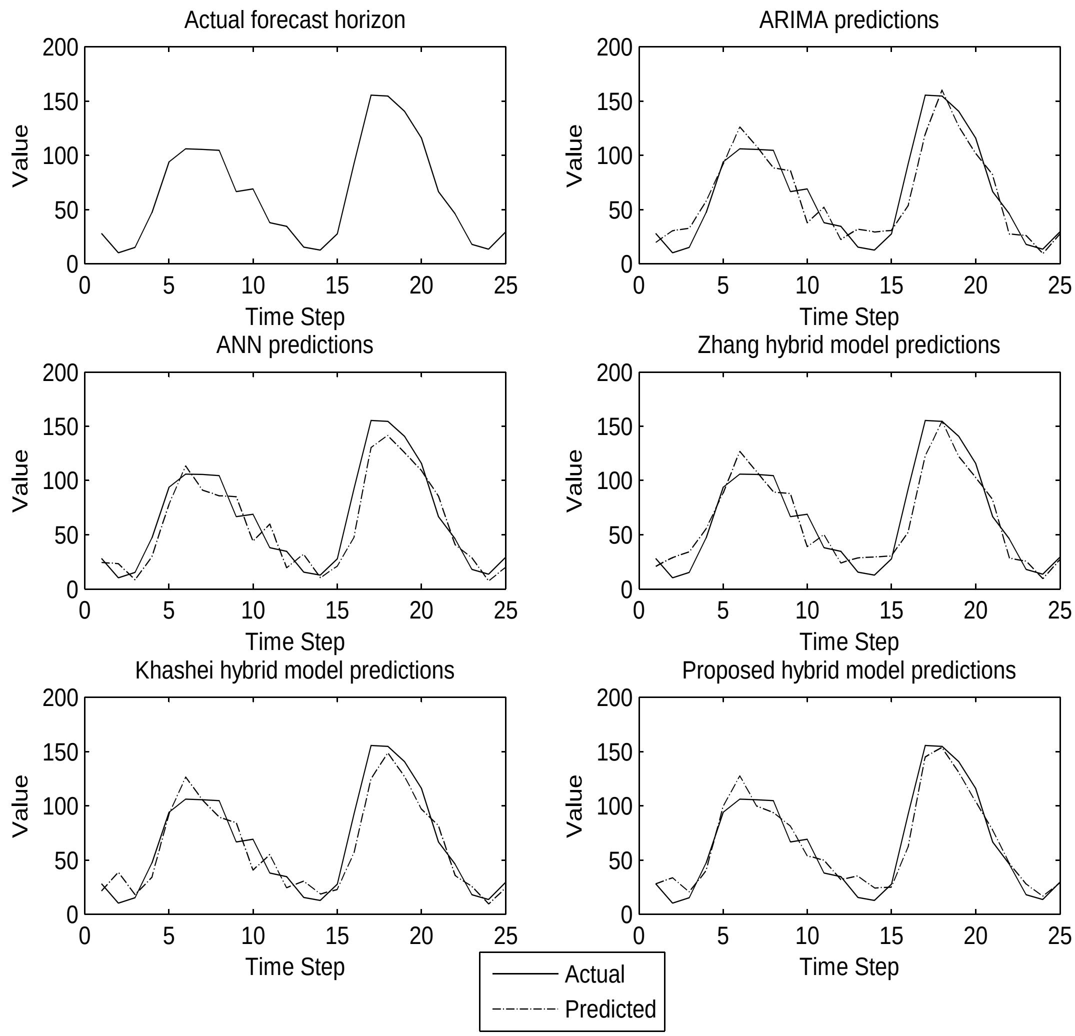

Figure 3.9: One-step-ahead predictions for sunspot data using various models

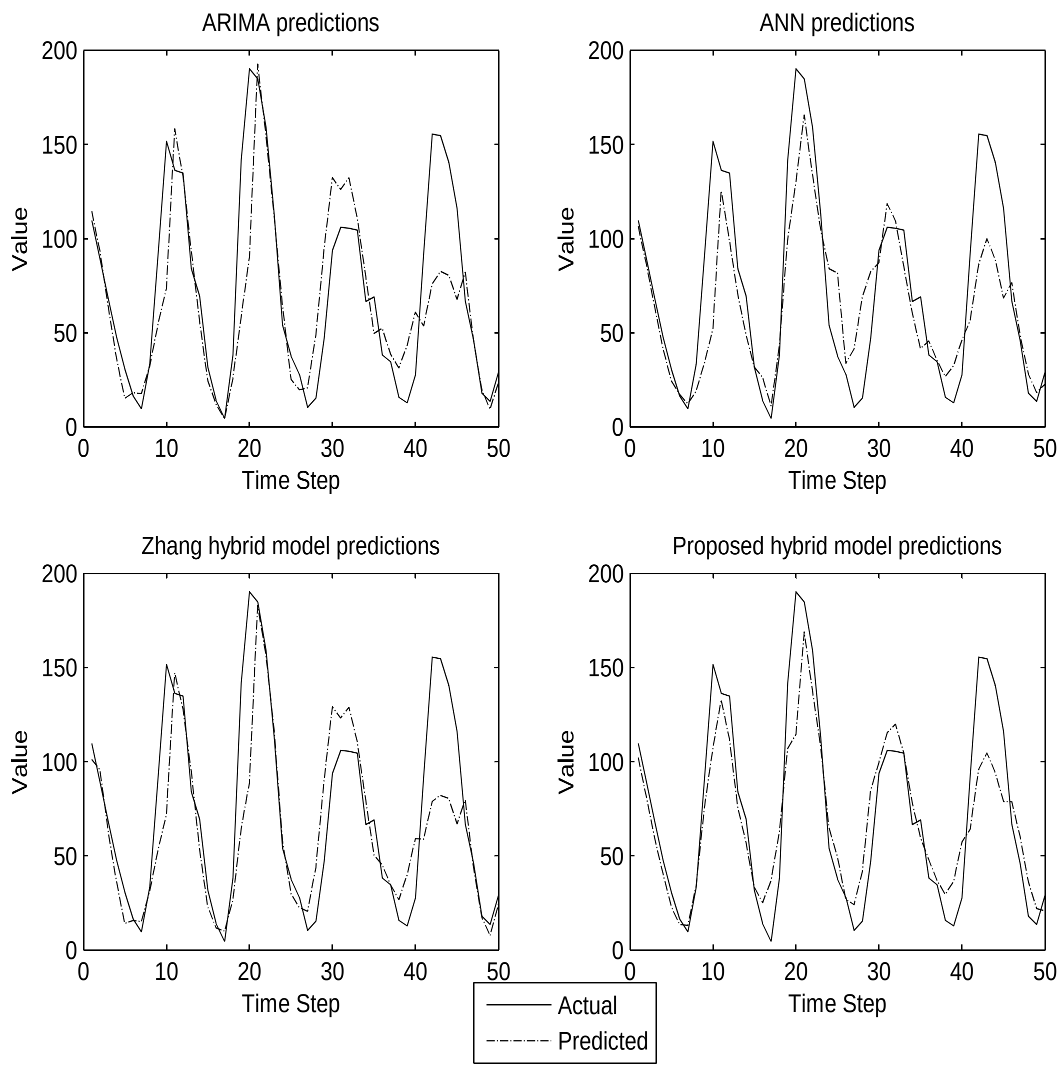

Figure 3.10: Five-step-ahead predictions for sunspot data using various models

3.3.3 Results for electricity price data

The electricity price data studied here were data for New South Wales from the Australian National Electricity Market [101] for the month of May in 2013. The data were available for every half hour. This was converted first to one-hour data, so that there were 24 data points for one day. So, for one month, 744 data points representing hourly electricity price data were taken as the given time series data set for forecasting. In one-step-ahead forecasting, the forecast horizon considered was 24 data points. Also, 24-step-ahead (one-day-ahead) forecasting was performed, where the forecast horizon was taken as 7 days, which means 168 data points. The ARIMA model fit in ARIMA, Zhang's hybrid and Khashei-Bijari hybrid model is ARIMA(1,0,1), whose model parameters are given in Table 3.5. In the proposed hybrid model, the ARIMA model fit is ARIMA$(1, 1, 1)$, whose model parameters are shown in Table 3.6. The prediction performance results for all of the models for both one-step-ahead and one-day-ahead forecasts are shown in Table 3.7. The original data set is shown in Figure 3.11. The one-step-ahead predictions are shown in Figure 3.12, and the one-day-ahead predictions in Figure 3.13. From the table and the figures, it can be seen that the proposed method outperforms the others for both one-step ahead and multi-step ahead forecasting.

When the proposed prediction model was used on the data, the data had a kurtosis of 28.4, indicating that the data were very highly volatile in nature. When these data were passed through an MA filter

Table 3.5: ARIMA$(1, 0, 1)$model parameters

Parameter	Value
Constant	62.062
AR1	0.436
MA1	0.102
Variance	284.5

Table 3.6: ARIMA$(1, 1, 1)$ model parameters

Parameter	Value
Constant	-0.0027
AR1	0.4476
MA1	0.006
Variance	1.185

of length 25, the trend component had a kurtosis of 3 and the residual component had a kurtosis of 18.2. An ARIMA model was fit to the trend component and an ANN was fit to the residual component. Note that when an ARIMA model is fit according to either Zhang's or Khashei and Bijari's model, the order of the model is same. But in the proposed method, the order of the ARIMA model is different. For example, in this case, for Zhang's and Khashei and Bijari's models, the ARIMA model used was ARIMA$(1, 0, 1)$, but for the proposed method the ARIMA model was ARIMA$(1, 1, 1)$. This was because when the trend component was separated, the ARIMA model fit to the data was entirely different from the ARIMA model fit directly to the data. From the results, it can be

Table 3.7: Performance comparison for electricity price data

	One-step-ahead		24-step (one-day) ahead	
	MAE	MSE	MAE	MSE
ARIMA	4.8233	36.4544	8.6682	121.5435
ANN	3.7374	22.4304	8.4138	109.8774
Zhang model	3.9204	27.0377	7.9437	107.3933
Khashei and Bijari model	3.8346	26.1396	NA	NA
Proposed model	**3.2342**	**18.2793**	**5.3219**	**53.0071**

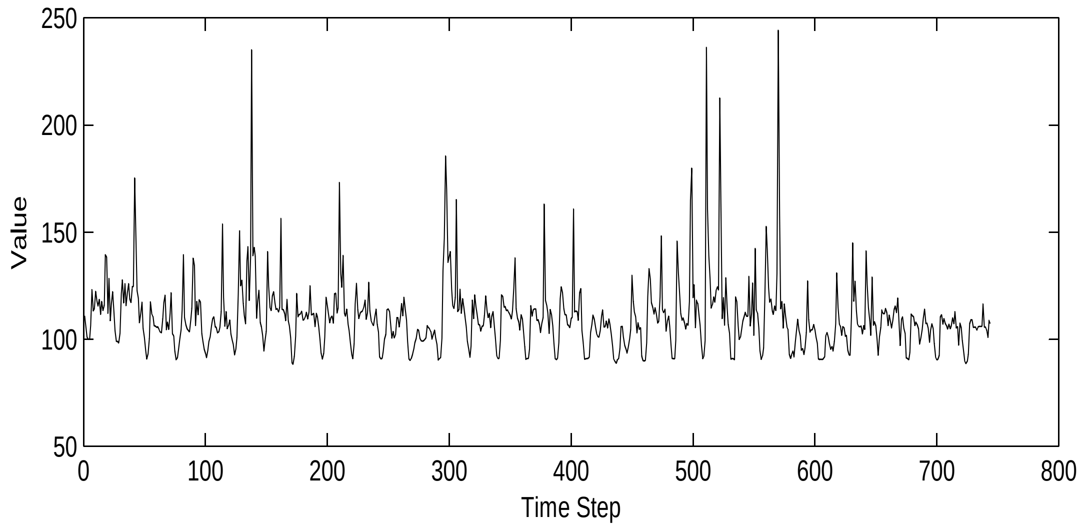

Figure 3.11: Electricity price time series data

concluded that the proposed method outperforms the other models discussed in this research work.

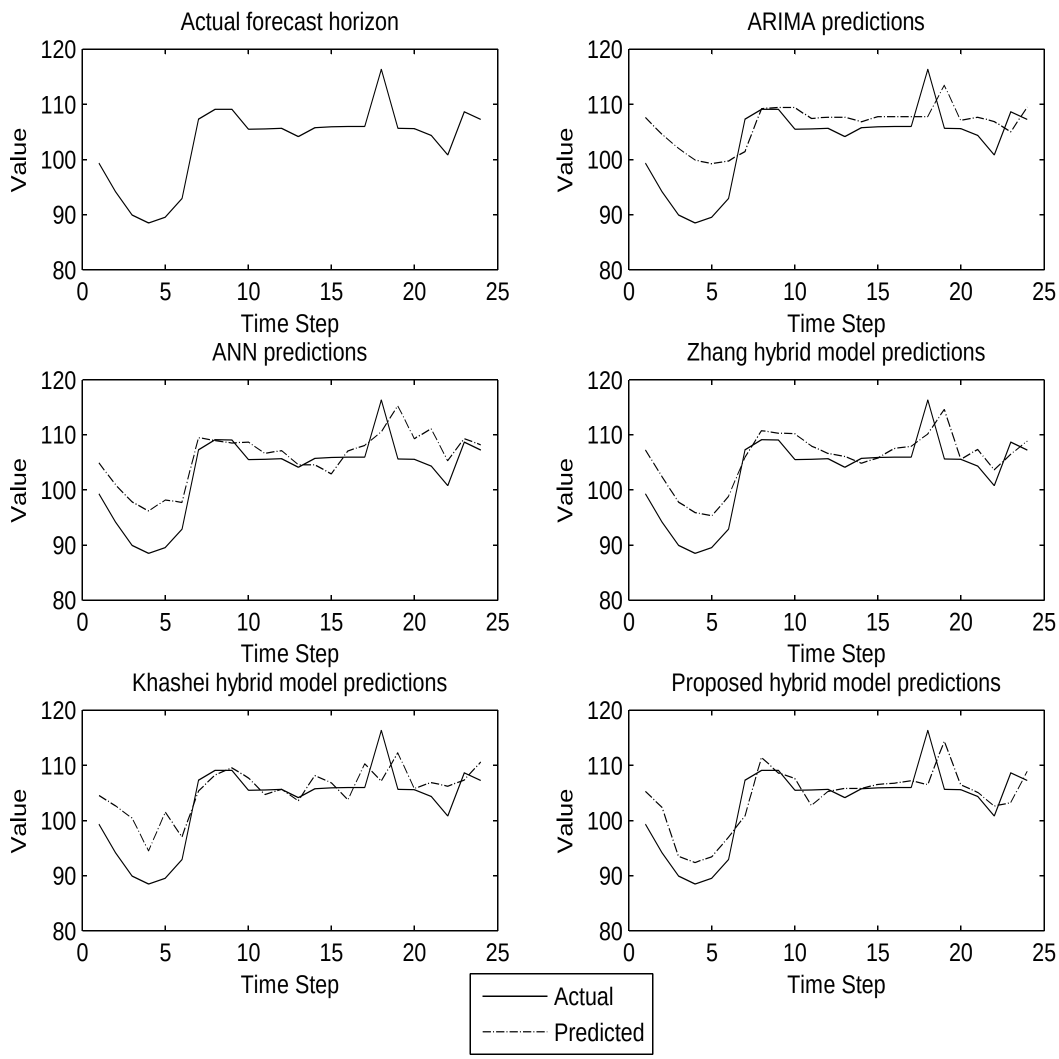

Figure 3.12: One-step-ahead predictions for electricity price data using various models

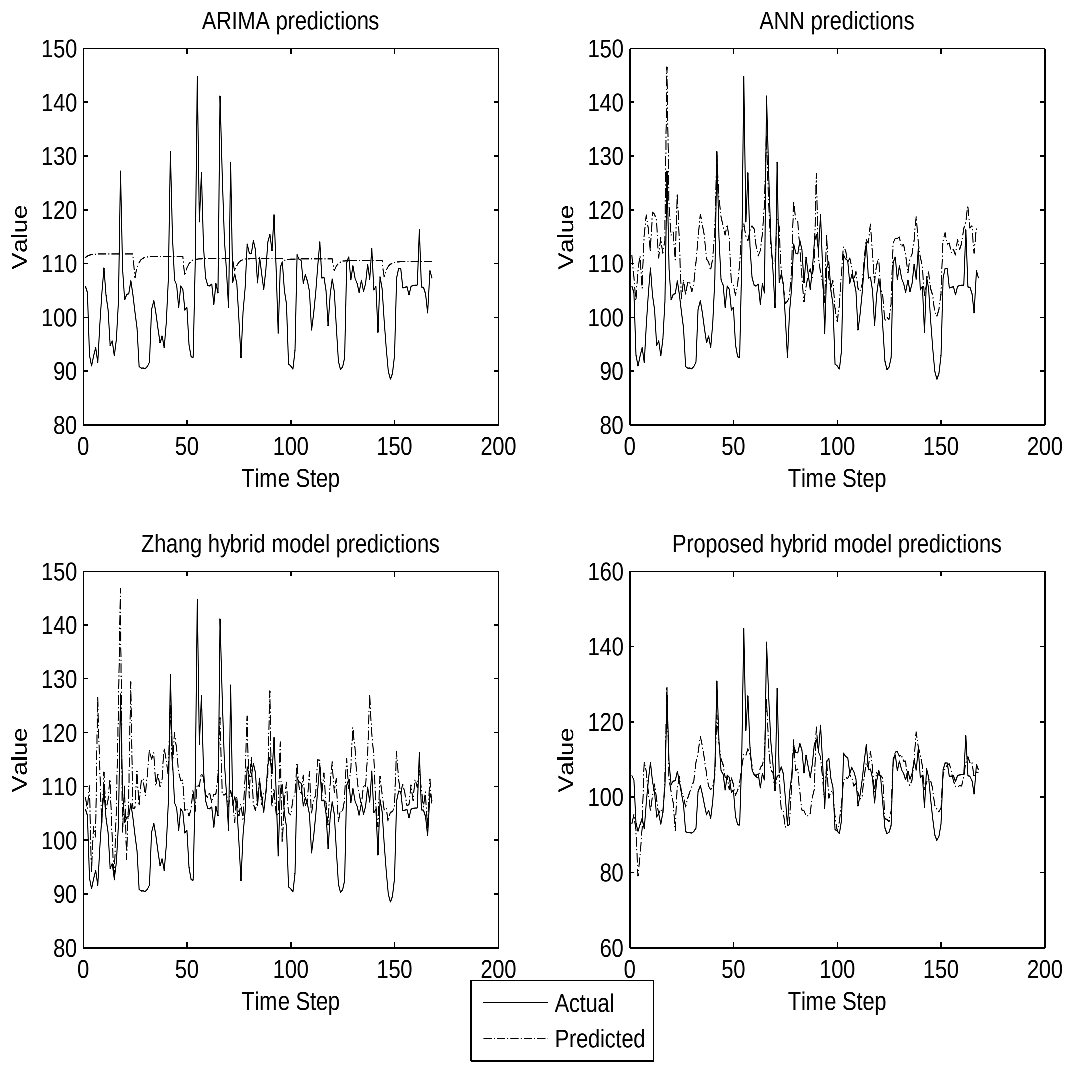

Figure 3.13: Five-step-ahead predictions for electricity price data using various models

Table 3.8: Performance comparison for L&T stock time series data

	One-step-ahead		Three-step-ahead	
	MAE	MSE	MAE	MSE
ARIMA	17.4141	629.7310	26.8604	1.5685
ANN	14.8967	428.7039	22.8979	1.0514
Zhang model	14.7062	389.8686	18.9346	0.9732
Khashei and Bijari model	14.8407	401.2017	NA	NA
Proposed model	**12.8226**	**261.6390**	**15.9454**	**0.7422**

3.3.4 Results for stock market data

The close prices of the Larsen and Turbo (L&T) company stock for 200 trading days before May 31, 2013 were chosen as the time series data for study. The data set was taken from [102]. For one-step-ahead prediction, the forecast horizon was chosen as 20 data points. Three-step-ahead forecasting was also performed, for which the forecast horizon was taken as 21 data points. The ARIMA model fit in ARIMA, Zhang's hybrid and Khashei-Bijari hybrid model is ARIMA(0,1,0). In the proposed hybrid model, the ARIMA model fit is ARIMA$(3, 2, 0)$, whose model parameters are shown in Table 3.9. The five models discussed in this research work were applied to these time series data, and the prediction performance results are tabulated in Table 3.8. The original data are shown in Figure 3.14. The one-step-ahead predictions are shown in Figure 3.15, and the three-step-ahead predictions in Figure 3.16. From the table and the results, it can be seen that the proposed method outperforms the other models. In the case of one-step-ahead prediction, the performance of Zhang's and Khashei and Bijari's models showed a very small improvement, whereas the proposed model showed significant improvement compared with Zhang's model. Also, in this case, the

Table 3.9: ARIMA$(3, 2, 0)$model parameters

Parameter	Value
Constant	0.0372
AR1	0.226
AR2	-0.166
AR3	0.476
Variance	0.560

accuracy of Zhang's model was better than that of Khashei and Bijari's model for one-step-ahead prediction. But the accuracy of the proposed method was much better than that of the others in terms of both MSE and MAE, as can be seen from the table.

The MA filter length in the proposed method was 90. The given time series data had a kurtosis of 2.04. The trend component of the MA filter had a kurtosis of 3, and the kurtosis of the residual component was 2. It can be seen that the original data had a kurtosis less than 3, so it was considered as highly volatile. Whenever the kurtosis is not 3, the data are highly volatile. If the kurtosis is greater than 3, the data are both highly outlier-prone and highly volatile. If the kurtosis is less than 3, the data are still highly volatile but less outlier-prone [103]. The data set considered in this subsection was less outlier-prone, whereas the sunspot and electricity price data were highly outlier-prone. Irrespective of whether or not the data were highly outlier-prone, the proposed model outperformed the other models, as seen from all of the results.

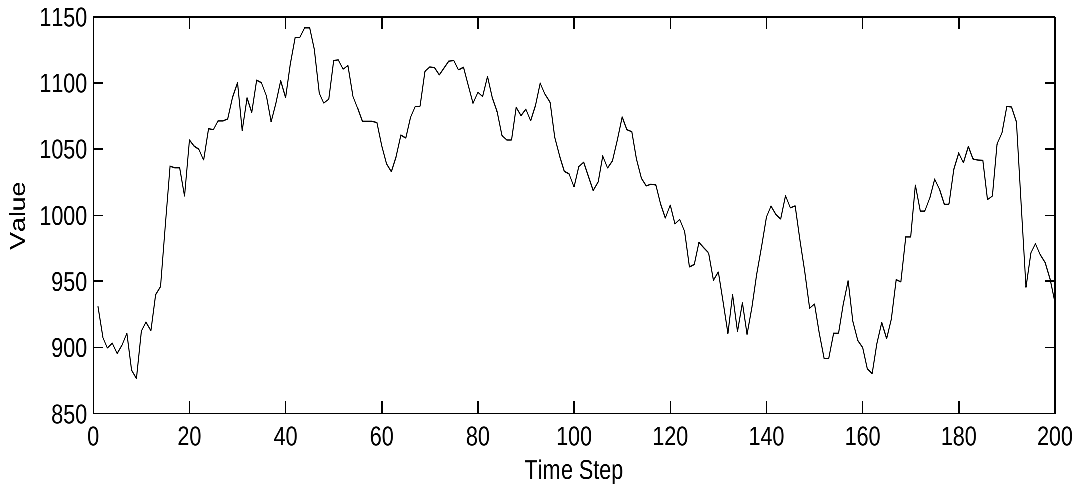

Figure 3.14: L&T stock market time series data

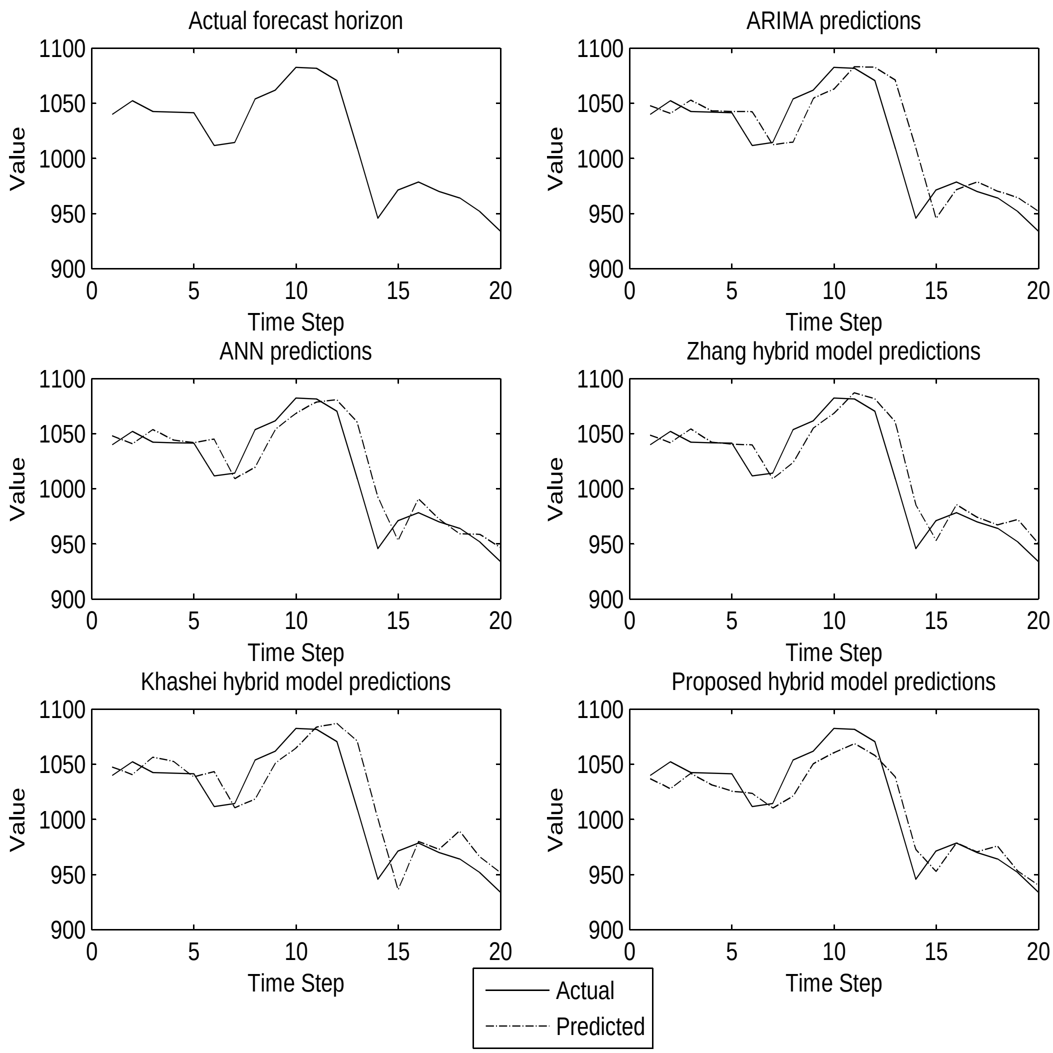

Figure 3.15: One-step-ahead predictions for financial data using various models

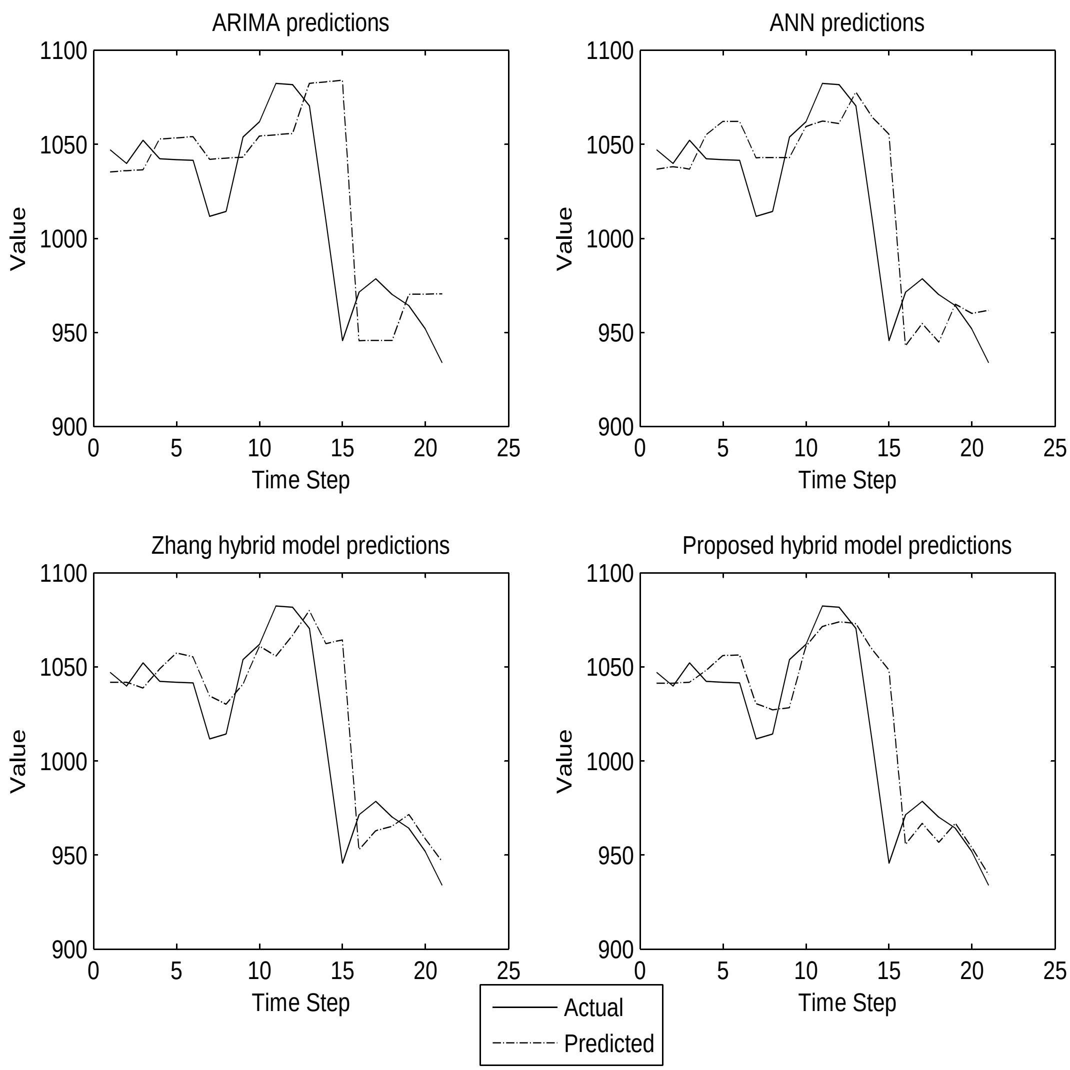

Figure 3.16: Five-step-ahead predictions for financial data using various models

3.4 Summary

Time series data originating from various applications, in general, comprise both linear and nonlinear variations. Linear ARIMA models and nonlinear ANN models cannot individually model such data accurately. Hybrid models which combine the strengths of ARIMA and ANN models are better than the individual types of models, as they are capable of exploiting the advantages of both types of models simultaneously. In this regard, this chapter presented a hybrid ARIMA-ANN based prediction model which is proposed using the statistical properties of ARIMA sequences for accurate prediction of TSD values. The model uses MA filter to decompose the given TSD into two data sets. Then ARIMA and ANN models are applied suitably to these decompositions. The forecasts from the hybrid model are obtained by adding the forecasts from the two individual models. This hybrid model is capable of both one-step ahead and multi-step ahead prediction. The model was applied to simulated time series data and to three available data sets of different kinds, namely sunspot data, electricity price data, and financial data. For both one-step ahead and multi-step ahead prediction, the proposed hybrid model has higher prediction accuracy in terms of MAE and MSE than several other models, such as ARIMA and ANN models and some existing hybrid ARIMA-ANN models of Zhang, Khashei-Bijari. Thus the hybrid model proposed in this chapter becomes a simple and accurate prediction model in many applications.