CHAPTER 6

APPLICATION OF VARIOUS PREDICTIVE MODELS ON SELECTED TSD

Chapter 6

APPLICATION OF VARIOUS PREDICTIVE MODELS ON SELECTED TSD

This chapter presents the application of ARIMA, ANN, GARCH and variants of ARIMA models like trend-ARIMA or wavelet-ARIMA model to some selected environmental TSD like average global temperature and monthly rainfall in India. It also presents the similarity of ARIMA model and adaptive filter based predictive model for prediction of TSD, which is validated using financial data originating from Indian stock market. In each case, the detailed results of modeling and prediction are presented.

The chapter is organized as follows. In section 6.1, the detailed results for application of ARIMA, GARCH, ANN, trend-ARIMA and wavelet-ARIMA models to average global temperature TSD is presented. In section 6.2, the detailed results for application of ARIMA, GARCH, ANN and wavelet-ARIMA models to rainfall of India TSD is presented. In section 6.3, the details of similarity in application of ARIMA and adaptive filter based prediction model are presented. The chapter summary is presented in section 6.4.

6.1 Prediction of average global temperature TSD

The average global temperature TSD from 1880 to 2010 is considered and prediction is performed using ARIMA, trend-ARIMA, wavelet-ARIMA, GARCH and ANN models. The prediction performed in this case is multi-step ahead prediction. In particular using values of global temperature from 1880 to 2000, the values from 2001 to 2010 are predicted, thus carrying out a 10-step ahead prediction. The results of the prediction are compared using MAPE, MaxAPE, MAE and RMSE. The prediction horizon is chosen as 10 data points from 2001 to 2010. On this TSD, using ARIMA modeling steps detailed in Chapter 1, the model order is found out to be ARIMA$(5, 1, 8)$.Applying trend-ARIMA model, by decomposing the given TSD using a moving-average filter, the trend component is fit using ARIMA$(11, 0, 11)$ and residual or error or noise component is fit using ARIMA$(5, 0, 5)$ model.

The wavelet-ARIMA model is fit on the raw data after it has been decomposed into one approximate and three detailed time series components. Wavelet filter used here is db5. Each decomposed TSD component is estimated using ARIMA modeling steps and then predictions are obtained. Using GARCH modeling steps, a GARCH$(1, 1)$ model is fit and predictions are obtained. Similarly using LM training algorithm for ANN, the ANN model is fit on the TSD and the corresponding predictions are then obtained. The actual values and the predictions obtained with

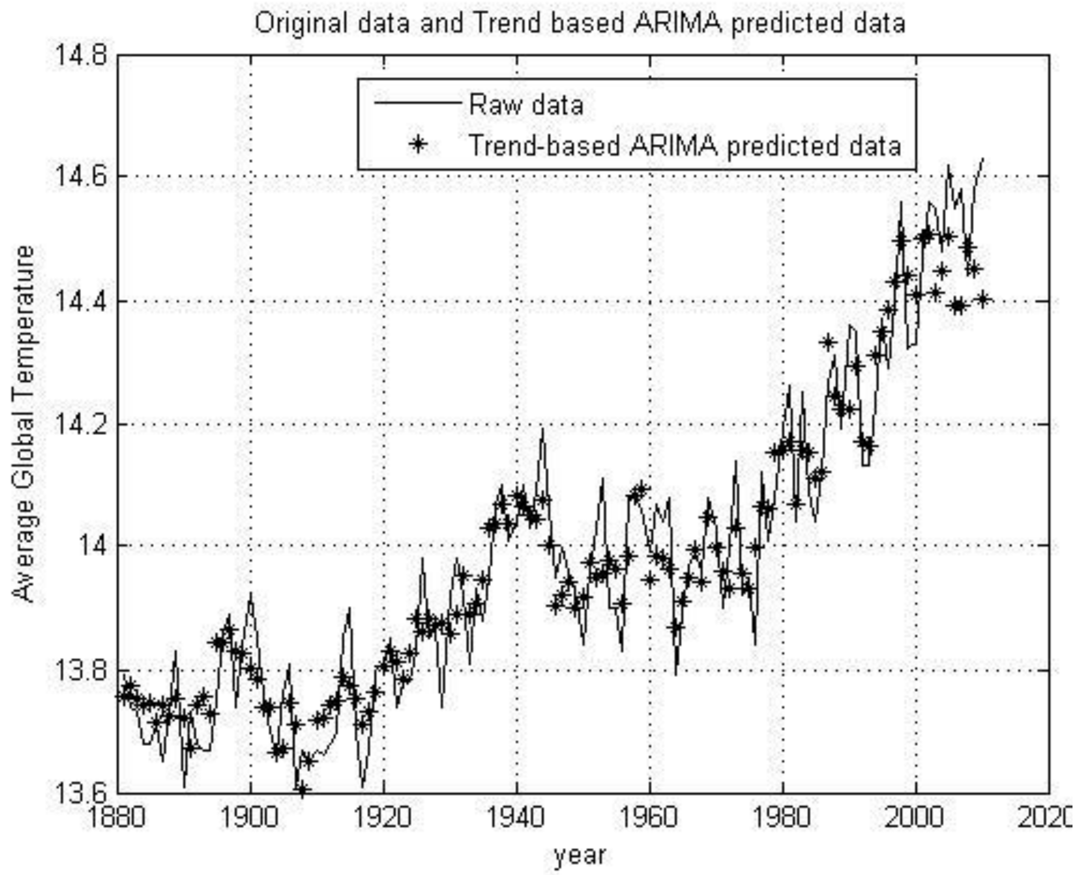

Figure 6.1: Raw data and Trend-based ARIMA predicted data.

all the above said models except trend-ARIMA model is shown in Figure 6.2. It can be observed that the data dynamics or data trend is preserved by the wavelet ARIMA model more accurately than the others. The prediction results of Trend-ARIMA model are given in Figure 6.1. The error performance measures are shown in Table 6.1. It can be observed that in this case Wavelet ARIMA model outperforms all the other models.

Table 6.1: Performance measures for Global Temperature

Method	MAPE %	MaxAPE %	MAE	RMSE
ARIMA	1.06	1.76	0.1545	0.1688
Trend-ARIMA	0.77	1.55	0.1119	0.1079
Wavelet ARIMA	**0.37**	**0.81**	**0.0542**	**0.063**
GARCH	1.8	3.67	0.2542	0.287
ANN	1.92	4.65	0.41	0.44

6.2 Prediction of rainfall TSD in India

The average monthly rainfall TSD from January 2010 to December 2010 has been considered for prediction using ARIMA, trend-ARIMA, wavelet-ARIMA, GARCH and ANN models. The prediction performed in this case is multi-step ahead prediction, which is in particular a 20-step ahead prediction. The data is identified less volatile TSD using the conditional standard deviation plot. The results of the prediction are compared using MAPE, MaxAPE, MAE and RMSE. On this TSD, using ARIMA modeling steps detailed in Chapter 1, the model order is found out to be ARIMA$(18, 0, 18)$.

The wavelet-ARIMA model is fit on the raw data after it has been decomposed into one approximate and three detailed time series components. Wavelet filter used here is db5. Each decomposed TSD component is estimated using ARIMA modeling steps and then predictions are obtained. Using GARCH modeling steps, a GARCH$(1, 1)$ model is fit and predictions are obtained. Similarly using LM training algorithm for ANN, the ANN model is fit on the TSD and the corresponding predictions are then obtained. The actual values and the predictions obtained with all

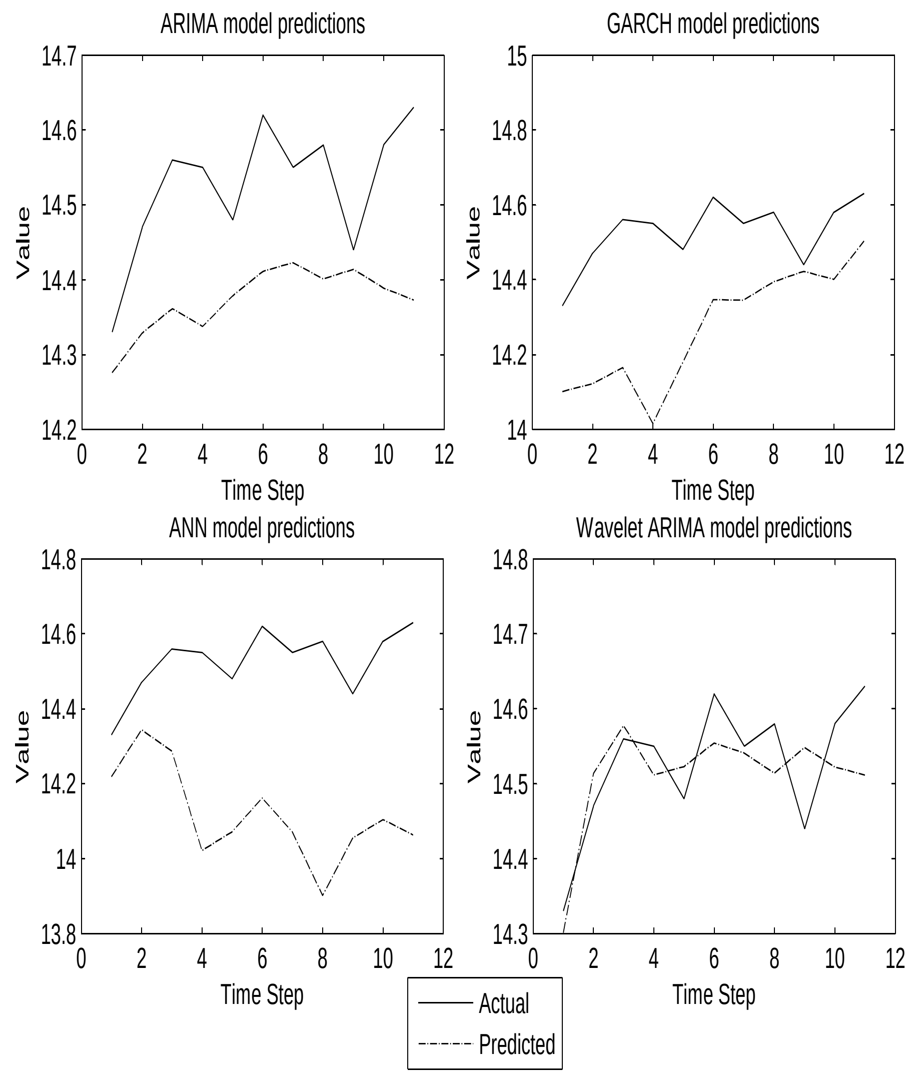

Figure 6.2: Forecast results for Average Global Temperature.

the above said models is shown in Figure 6.3. It can be observed that the data dynamics or data trend is preserved by the wavelet ARIMA model more accurately than the others. The error performance measures are shown in Table 6.2. It can be observed that in this case Wavelet ARIMA model outperforms all the other models except ANN. The performance obtained from wavelet-ARIMA model is almost similar to that given by ANN model.

Table 6.2: Performance measures of rainfall data

Method	MAPE%	MaxAPE%	MAE	RMSE
ARIMA	7.79	18.9	89.6222	105.48
Wavelet-ARIMA	**6.86**	**16.47**	**79.6636**	**93.38**
GARCH	9.69	30.93	111.2257	144.21
ANN	6.72	14.01	74.631	92.86

6.3 Prediction of stock market variables

To predict stock market TSD like close price of a stock, high price, low price, open price, number of trades and traded volume, in this section it is investigated if instead of ARIMA model which requires the laborious Box-Jenkins approach, if least mean squares (LMS) based adaptive filter can be used as a forecast model. So initially, the basics of an adaptive filter based prediction model is described. Then this model along with ARIMA model are applied on six different TSD of stock market, and it is verified that the adaptive filter based model indeed gives almost similar performance as that of the ARIMA model. Adaptive time series filters have been used to forecast economic variables in [111].

6.3.1 Adaptive filter based prediction model

Adaptive filters are most commonly encountered in signal processing and communications. These can be used not only for forecasting purpose but can also be used for various other applications like system identification (for seismic data), noise or interference cancelation, inverse modeling etc. Such a filter is used for prediction purpose in this section.

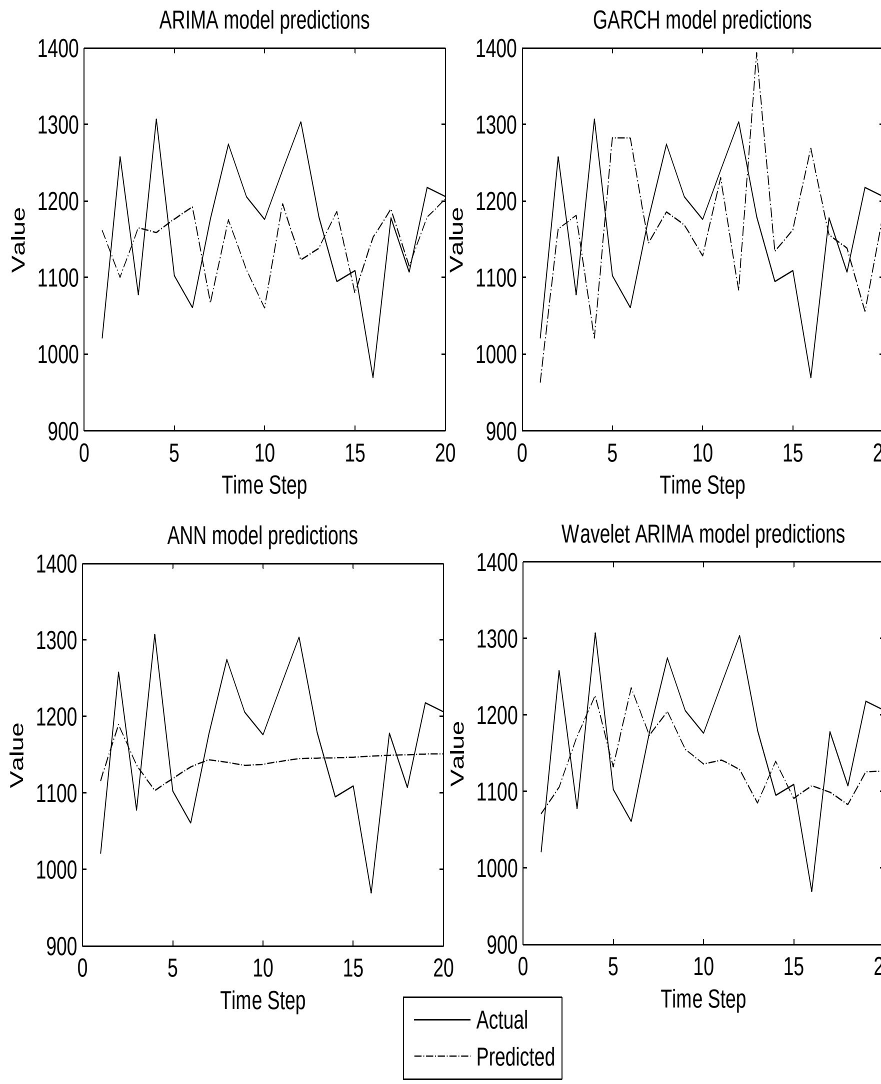

Figure 6.3: Forecast results for rainfall data.

A predictive model using an Adaptive filter is shown in Figure 6.4. Adaptive filter is a model in which the output depends on algebraic sum of previous inputs and the filter coefficients adapt (vary) as a function of time. These coefficients are adapted based on the error value, where error is the difference between actual and the estimated values. If the given TSD is purely of AR type, this filter model fits the TSD the best. But if the given TSD has MA component, the required order of AR can be made very high such that it fits the MA type TSD in a better way [100].

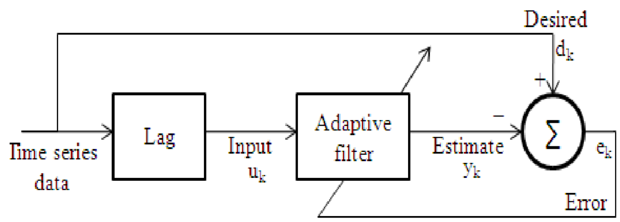

Figure 6.4: Adaptive filter model block

The advantage is that the adaptive filter model is a linear model and does not require any non-linear solution unlike Box-Jenkins approach. As shown in Figure 6.4, the adaptive filter block requires 'desired' and 'input' data. For a given TSD the present value to be estimated, becomes the desired data. TSD lagged by a definite order becomes the input value to the filter. The filter output is the estimated present value. The error which is the difference between actual and estimated values is calculated at every instant. Then the filter coefficients are varied accordingly, in a direction to decrease the error. After the end of training period, the filter coefficients are fixed. Then based on the input to the filter and filter coefficients the future data forecasts are obtained. For this adaptive filter model to be fit on the data, the given data need not be made stationary unlike ARIMA approach where stationarity is of prime concern. The adaption equation can be framed based on various algorithms like Steepest Descent, Least Mean Squares (LMS), and Recursive Least Squares (RLS) etc. The LMS adaptation equation is used here, which is given as in 6.1 and described next.

In (6.1), e_k represents error value, d_k represents desired data value,

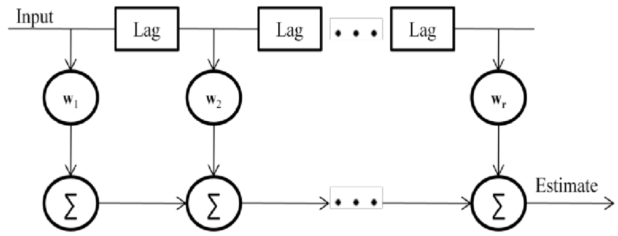

Figure 6.5: LMS based aptive model block

u_k represents input data vector to adaptive filter block, W_k represents filter coefficient vector, in the k^{th} iteration. y_k represents estimated output data of the filter. μ indicates adaption step size used in LMS algorithm. The adaptive filter model can have varying number of filter coefficients. The filter coefficients are also called as taps or weights of the filter. If the number of taps is four, then length of the filter is four. The number of taps in an adaptive filter is generally chosen for minimum error. But this criterion may require large number of taps, which may not be feasible. So the number of taps is a tradeoff between error and feasibility. Note that d_k and e_k are values whereas u_k and W_k are vectors. u_k is formed by arranging the inputs of Adaptive filter at each tap point in k^{th} iteration, as a vector. W_k is formed by arranging the tap values of the filter in k^{th} iteration, as a vector.y_k is obtained by multiplying transpose of weight vector W_k with the input data vector u_k. $(.)^H$ in (6.1) represents the Hermitian transpose of the vector $(.)$. The adaptive filter block with these above mentioned notations is shown in Figure 6.5. The model equation for prediction after the weights are found out is detailed again in (6.2), which is same as the last equation in (6.1).

$$\begin{aligned} W_{k+1} &= W_k + \mu u_k * e_k \\ e_k &= d_k - y_k, \\ y_k &= W_k^H u_k \end{aligned} \tag{6.1}$$

$$y_t = W_1 y_{t-1} + W_2 y_{t-2} ...... + W_p y_{t-p} + e_t \tag{6.2}$$

The Adaptive filter coefficients are set to zero initially. While training progresses the weights are updated using adaption equations given in (6.1). After the training is complete, the filter coefficients are fixed and estimated data is obtained. By changing the step size parameter and the number of taps, an optimal model is obtained. This implementation is done in MATLAB. The comparison of ARIMA and adaptive estimates using performance measures MAPE and MaxAPE is also done in MATLAB.

6.3.2 Prediction results:

The forecasting is done on the variables listed in Table 6.3. This is a one-step ahead prediction. Normalization as a preprocessing step is performed on the dataset shown in Figure 6.6. This data is given as input to fit ARIMA model. The model order for each of the stock variable shown in Table 6.3 is shown in Figure 6.7.

Preprocessed actual data is given as input to adaptive filter model shown in Figure 6.4, which is implemented in MATLAB. Lag operation is

Table 6.3: Stock variables

SNo	Variable Name
1	Open Price
2	High Price
3	Low Price
4	Close Price
5	Number of shares
6	Number of trades
7	Turnover

performed on the actual TSD and the resultant data is given as input to adaptive filter, u_k. Desired data, d_k, is the actual data itself, without any lags. The estimate y_k is compared with d_k and based on error value, the adaptive filter coefficients are changed. At the end of training period the filter coefficients are obtained. Then output for the prediction interval is computed, which is compared with predicted ARIMA values and also with the actual data. The results are shown in Figure 6.8 to Figure 6.14. It can be observed that the first four stock market variables have less number of outliers where as the last three variables have large number of outliers.

Performance of both these models is tabulated in Table 6.4. As seen from the Table, MAPE is low with ARIMA model, for those variables which have less number of outliers; MAPE is low with adaptive filter model for those variables with large number of outliers. So adaptive model outperforms ARIMA model, in case TSD has large number of outliers and ARIMA model outperforms adaptive filter model, in case TSD has less number of outliers.

For high price and number of shares TSD, the actual data, ARIMA

Date	Open Price	High Price	Low Price	Close Price	No.of Shares	No. of Trades	Turnover (Rs.)
1-Feb-07	465	471.55	451.1	457.75	4451227	52216	2038163503
2-Feb-07	460	472.4	459.6	462.95	3565172	31843	1662684737
5-Feb-07	465	472.05	455.05	469.9	1310441	23589	609167209
6-Feb-07	474.9	475.5	463.55	465.2	1042249	15947	488922665
7-Feb-07	467	469.4	461.2	464.3	870061	12860	404741805
8-Feb-07	465.9	465.9	458.2	461.6	926219	11368	428022703
9-Feb-07	468.5	470	451.5	453.25	788889	13922	359770806
12-Feb-07	455.8	458.2	442.55	443.75	1788184	20476	802308405
13-Feb-07	447	448	430	432.25	1317505	23544	577372019
14-Feb-07	434.9	442.9	427.55	438.9	2333596	20014	1020017084
15-Feb-07	445	446.9	440.65	442.1	824839	9905	365642892
19-Feb-07	445.7	447	437.6	443.85	741303	9311	328449638
20-Feb-07	446	452	443.3	444.15	757573	9892	338714616
21-Feb-07	446.8	458.4	441.5	455.2	1109785	15459	500762035

Figure 6.6: Stock data sample

Model Description

			Model Type
Model ID	Open Price	Model_1	ARIMA(1,1,7)

Model Description

			Model Type
Model ID	High Price	Model_1	ARIMA(1,1,0)

Model Description

			Model Type
Model ID	Low Price	Model_1	ARIMA(0,1,8)

Model Description

			Model Type
Model ID	Close Price	Model_1	ARIMA(2,1,2)

Model Description

			Model Type
Model ID	No.of Shares	Model_1	ARIMA(0,1,7)

Model Description

			Model Type
Model ID	No. of Trades	Model_1	ARIMA(0,1,3)

Model Description

			Model Type
Model ID	Total Turnover (Rs.)	Model_1	ARIMA(0,1,7)

Figure 6.7: Model orders for stock variables

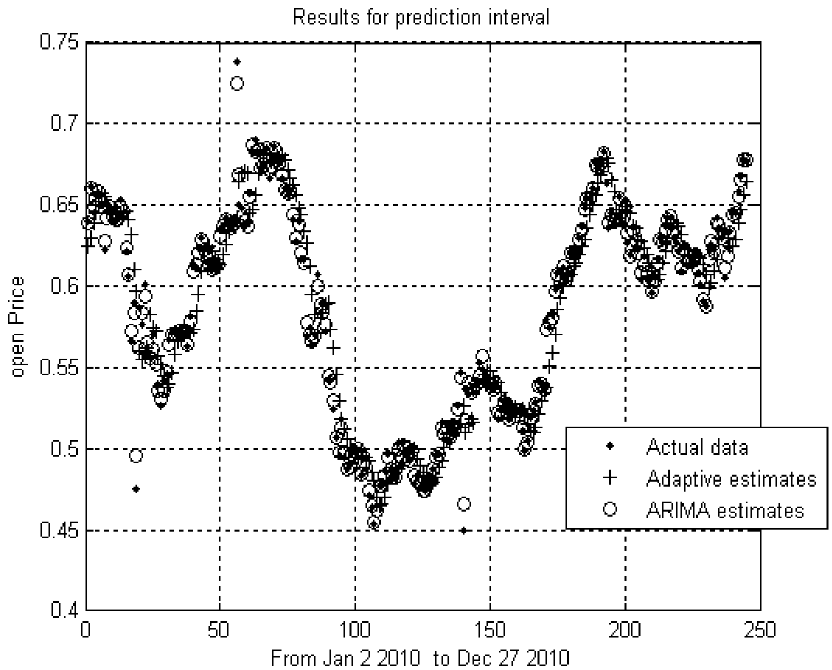

Figure 6.8: Prediction results using ARIMA and adaptive filter models for open price TSD

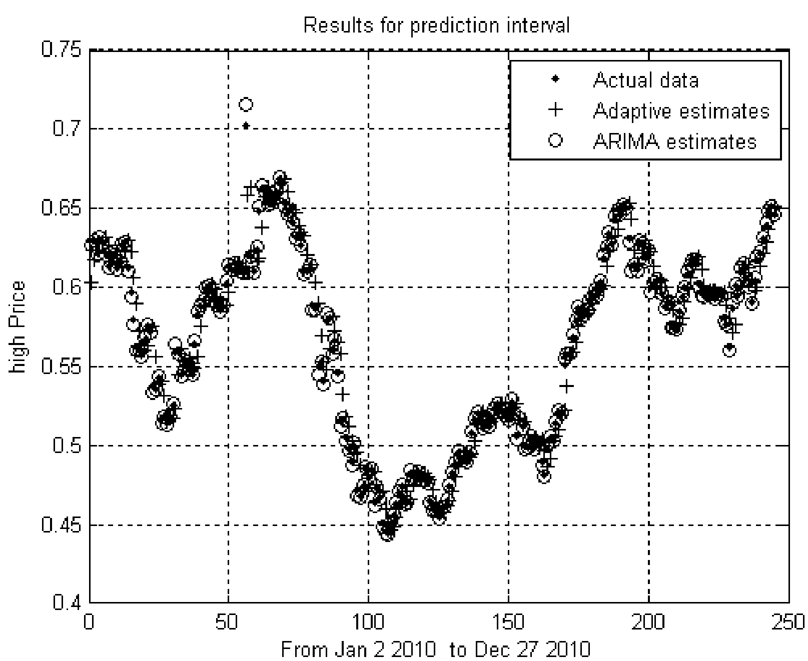

Figure 6.9: Prediction results using ARIMA and adaptive filter models for high price TSD

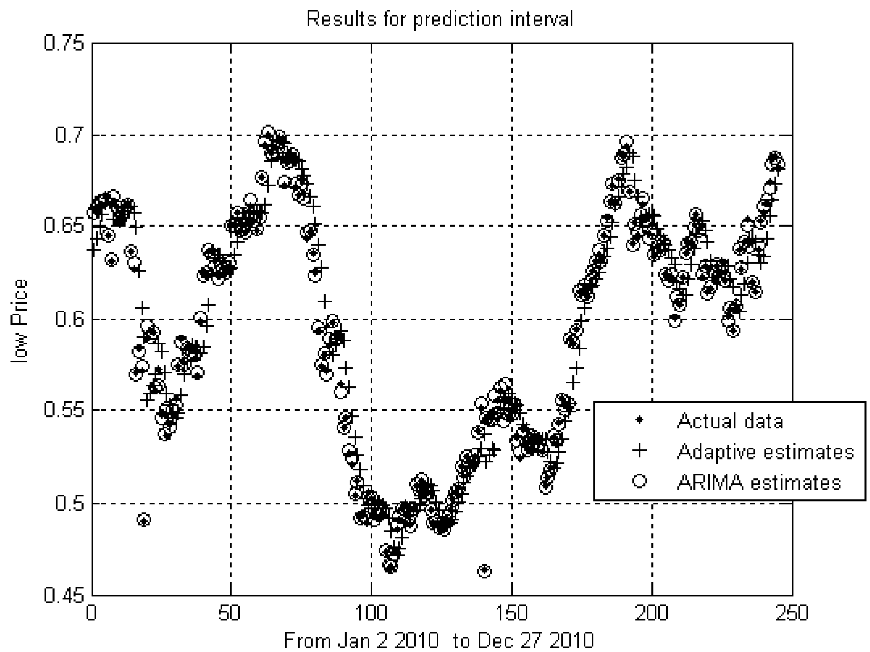

Figure 6.10: Prediction results using ARIMA and adaptive filter models for low price TSD

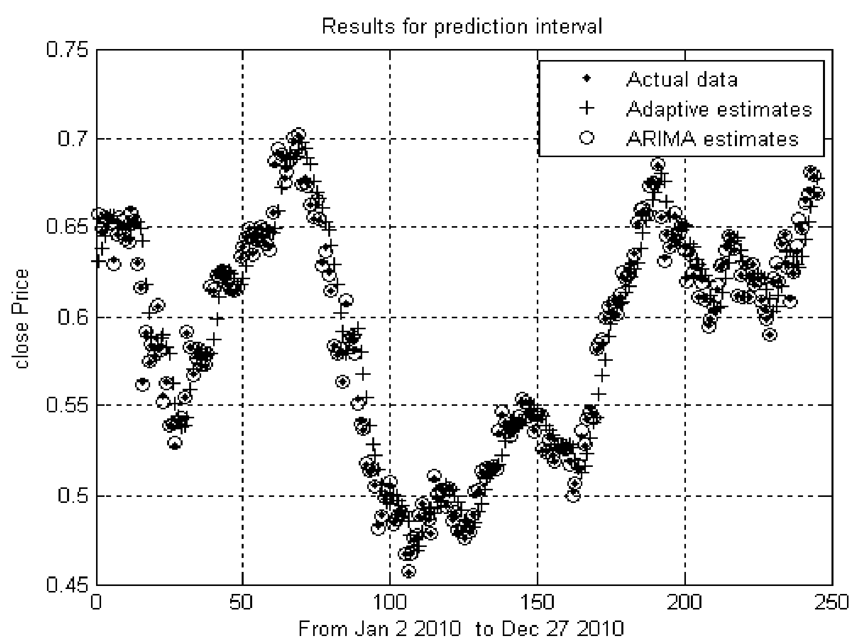

Figure 6.11: Prediction results using ARIMA and adaptive filter models for close price TSD

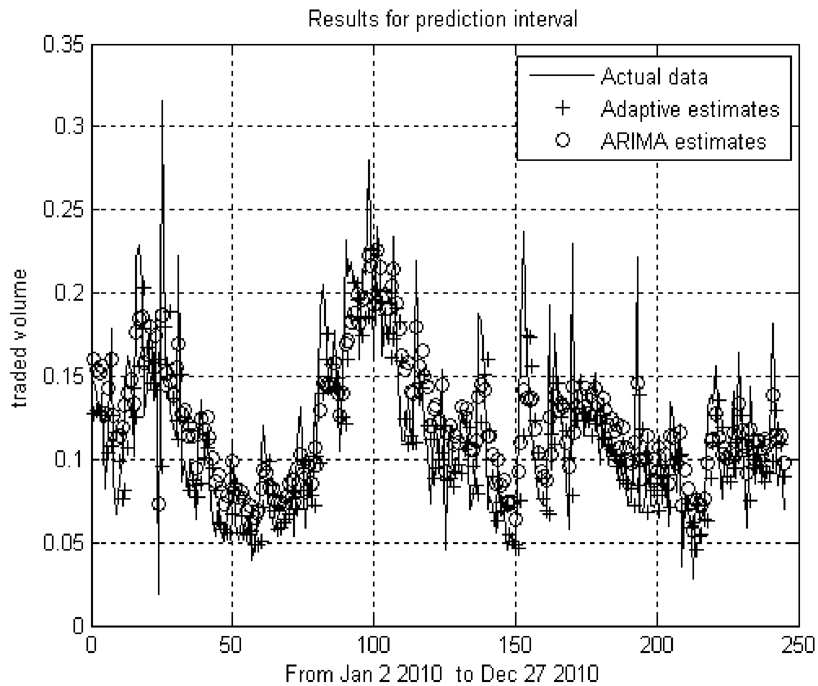

Figure 6.12: Prediction results using ARIMA and adaptive filter models for number of shares TSD

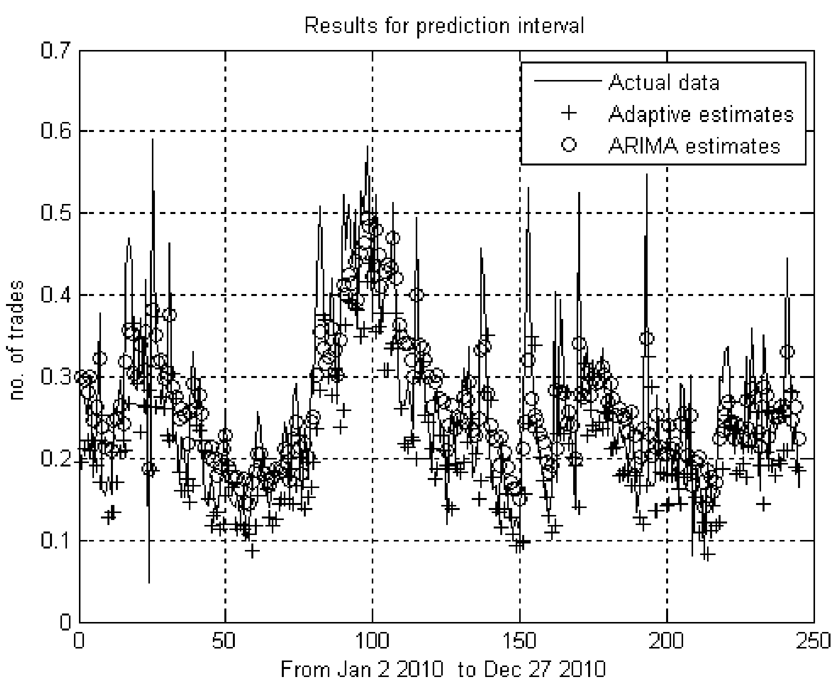

Figure 6.13: Prediction results using ARIMA and adaptive filter models for number of trades TSD

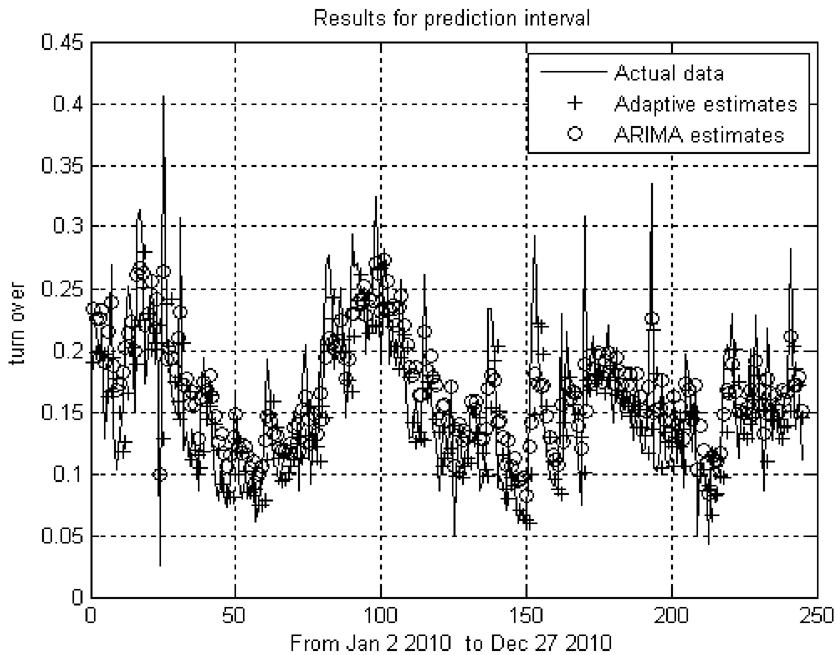

Figure 6.14: Prediction results using ARIMA and adaptive filter models for turnover TSD

estimated data and Adaptive estimated data for the complete training and prediction periods is shown in Figure 6.15 and Figure 6.16 respectively for comparison. High price data has less number of outliers whereas Traded Volume (Number of Shares) data has more number of outliers as seen from the figures. The number of taps for the adaptive filter is optimized and the value ranges between 3 and 5 for all the variables under consideration. The value of step size parameter is chosen to be small which ranges between, 0.01 to 0.08, depending on the variable under consideration.

Table 6.4: Performance measures

SNo	Variable Name	Adaptive MAPE	Adaptive MaxAPE	ARIMA MAPE	ARIMA MaxAPE
1	Open Price	0.0324	0.2845	**0.0210**	**0.2294**
2	High Price	0.0257	0.1271	**0.0148**	**0.1524**
3	Low Price	0.0324	0.2378	**0.0180**	**0.1947**
4	Close Price	0.0300	0.1523	**0.0171**	**0.0924**
5	Number of shares	**0.3318**	**6.3883**	0.3515	8.0813
6	Number of trades	**0.3055**	**4.3596**	0.3000	6.2514
7	Turnover	0.3325	**6.8371**	**0.3473**	8.4461

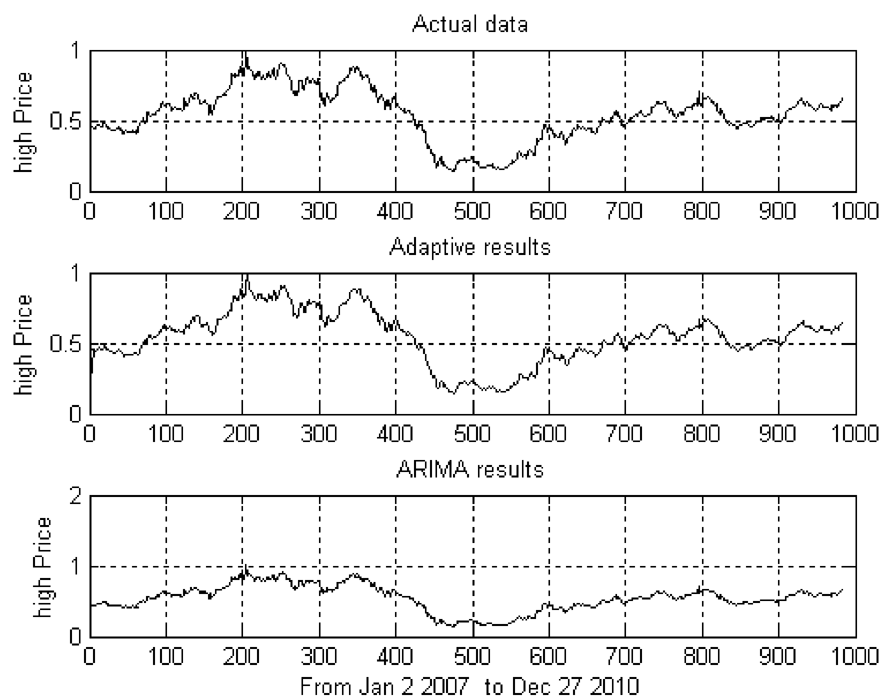

Figure 6.15: High price for training and prediction periods

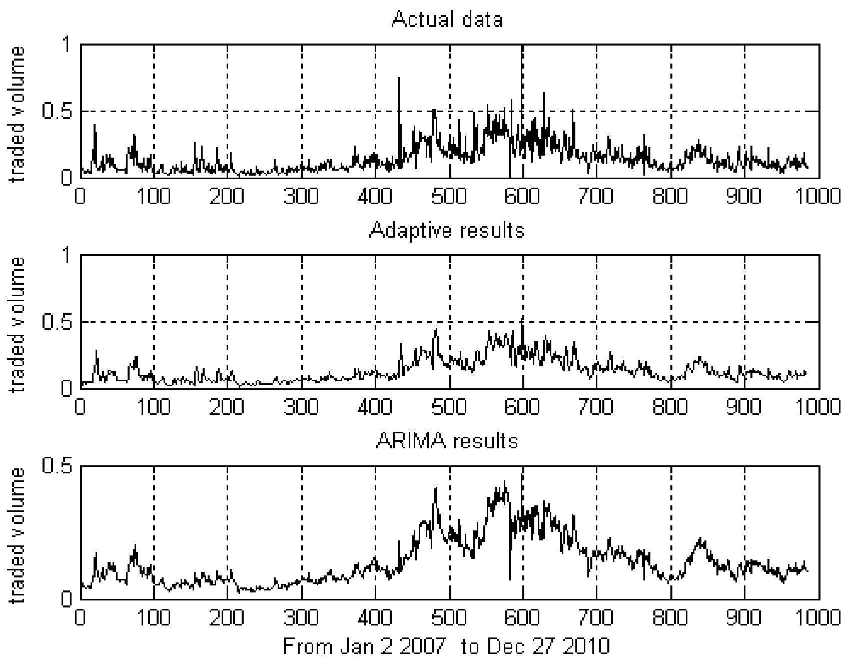

Figure 6.16: Number of shares for training and prediction periods

6.4 Summary

In this chapter, for the prediction of global temperature, the use of ARIMA, GARCH, ANN models along with trend-ARIMA and wavelet-ARIMA model are investigated. It is observed that the wavelet-ARIMA suits this TSD better than the other models for a 10-step ahead forecast. For the prediction of monthly rainfall data in India, the application of four models ARIMA, GARCH, ANN and wavelet-ARIMA are studied, the results of which showed that the error performance of wavelet-ARIMA and ANN are almost same, but wavelet-ARIMA retains the data trend better than the ANN and other models, for a 11-step ahead prediction and hence is more apt compared to the other models.