Time Series Analysis and Forecasting of Temperatures in the Sylhet Division of Bangladesh

A.H. Nury¹*, M. Koch² and M.J.B. Alam³

¹ Department of Civil Engineering, Leading University, Sylhet-3100, Bangladesh
² Department of Geohydraulics and Engineering Hydrology, Kassel University, Kassel, Germany
³ Department of Civil and Environmental Engineering, Shahjalal University, Sylhet, Bangladesh
*e-mail: hasancee@yahoo.com

Abstract

Time series analysis and forecasting has become a major tool in numerous applications in meteorology and other environmental areas to understand phenomena, like rainfall, humidity, temperature, draught etc. Here ARIMA (Auto Regressive Integrated Moving Average) models have been set-up and used to carry out short-term predictions of monthly maximum and minimum temperatures in the Sylhet and Maulvibazar districts in north-east Bangladesh. Using the classical Box-Jenkins methodology, stationary, seasonal ARIMA models for the temperatures recorded at two stations in Sylhet division between 1977 and 2011 are set up. Verification of the models has been done for the 2010-2012 time period. Based on the inspection of the ACF, PACF autocorrelation plots, the most appropriate orders of the ARIMA models are determined and evaluated using the AIC-criterion. For the maximum and minimum temperatures at Sylhet station ARIMA (1,1,1) (1,1,1)² and ARIMA (1,1,1) (0,1,1)², respectively, are obtained, whereas the respective models for the Maulvibazar station are ARIMA (1,1,0) (1,1,1)³ and ARIMA (0,1,1) (1,1,1)²³. Using these ARIMA-models one-month-ahead forecasts of the temperatures at the two stations for years 2010 and 2011 are carried out.

INTRODUCTION

Climate change is for many countries in the world one of the biggest environmental threats to food production, water availability, forest biodiversity and livelihoods [1]. Moreover, it is widely believed that developing countries in tropical regions of the world, e.g. Bangladesh, will be impacted more severely than developed ones [2][3].

Understanding the nature and scale of possible climate changes in north-eastern Bangladesh is of importance to the policy makers and people who are working there as it gives them a chance to be prepared for better mitigation and adaptation measures. For that purpose time series analysis of weather data can be a very valuable tool to investigate its variability pattern and, maybe, even to predict short- and long-term changes in the time series. Although any individual extreme climate event cannot be attributed unequivocally to climate change, the probability of high temperature events will increase if there is an underlying trend of rising mean temperature. In fact, according to data from the reinsurance industry, the number of climate-related disasters has increased significantly the 1970's [4].

Time series analysis and forecasting has become a major tool in numerous hydro-meteorological applications to study trends and variations in variables like rainfall, humidity, temperature, streamflow and many other environmental parameters [2][3][4][5]. A number of classical time series studies have been conducted in recent years to assess the nature of the climate change [1][2][3], as it has occurred over the world as well as in Bangladesh [4] in the recent past and as it will more in likely do so in the future.

The objectives of the present study are
- To build ARIMA models for the two temperature stations in the Sylhet division.
- To verify the selected ARIMA models
- To carry out short-term prediction for the temperatures at these stations

METHODOLOGY

Climate data:
Monthly temperature data covering the Sylhet and the neighboring Maulvibazar district has been collected from the Bangladesh Meteorological Department (BMD) which is the principal organization gathering meteorological data in Bangladesh. Temperature station of Sylhet district is situated in Sylhet at latitude 24°53'40", longitude 91°52'61" and Maulvibazar district is situated in Sreemongal upazila at latitude 24°18'31", longitude 91°43'49". The temperature data covers a period of 34 years, from 1977 to 2011. The study area is shown in the Figure 1.

ARIMA model:
The acronym ARIMA stands for ‘autoregressive
integrated moving average’ and they are sometimes also called Box-Jenkins models [6]. An autoregressive model of order \( p \) is conventionally classified as AR (\( p \)) and a moving average model with \( q \) terms is known as MA (\( q \)). A combined model that contains \( p \) AR-terms and \( q \) MA-terms is called an ARMA (\( p, q \)) model [6]. To make a generally non-stationary time series stationary time-shifted (by \( d \)-lags, whereby in most cases \( d=1 \)) differences are computed before further processing. Such a model is then classified as ARIMA (\( p, d, q \)), where the symbol “I” signifies “integrate”.

Assuming that the original data \( X_t \) has been made stationary by taking \( d \) nonseasonal differences (whereby in most cases \( d=1 \)), an ARMA (\( p, q \)) model for this new, stationary time series \( Y_t \) is as follows:

\[
Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots - \theta_q e_{t-q} \tag{1}
\]

where, \( c = \) constant term, \( \phi = \) autoregressive parameter, \( \theta = \) moving average parameter, \( e_t = \) the error term at time \( t \)

For climate data which usually follows a seasonal, i.e. an annual cycle, it is more appropriate to use a seasonal ARIMA (\( p, d, q \)) (\( P, D, Q \))\( S \) model, whereby \( P \) is the order of the seasonal AR-model; \( D \) is the order of the seasonal differencing (for monthly data, usually, \( D=12 \)) and \( Q \) is the order of the seasonal MA-model and \( s \) is the number of periods in the season (\( s=12 \), for an annual cycle) [7].

The general form of such a seasonal ARIMA (\( p, d, q \)) (\( P, D, Q \))\( S \) model, can be written in backshift notation as

\[
\phi_AR(B)\phi_SAR(B^s)(1-B)^d(1-B^s)Y_t = \theta_MA(B)\theta_SMA(B^s)e_t \ldots \ldots \tag{2}
\]

where \( \phi_AR = \) non-seasonal AR- parameter, \( \theta_MA = \) non-seasonal MA- parameter, \( \phi_SAR = \) seasonal AR-parameter, \( \theta_SMA = \) seasonal moving average parameter, \( B = \) backward shift operator.

To identify a perfect ARIMA model for a particular time series, Box and Jenkins [6] proposed a methodology that consists of four phases, namely; i) model identification; ii) estimation of the model parameters; iii) diagnostic checking for the identified model appropriateness and iv) application of the final model, i.e. forecasting. Further details on the various procedures involved in these four steps are provided in [6], [8].

As the original monthly temperature time series analyzed here are non-stationary, and have a 12-month seasonality, they are all differenced non-seasonally at lag \( d=1 \) and differenced seasonally at lag \( D=12 \).

The next step in the Box and Jenkins methodology [6] consists then in the determination of the orders \( p \) and \( q \) in ARMA\( p, q \) model. This is done by examination of the partial autocorrelation plot and
the partial autocorrelation plot of the time series, respectively. ACF- and PACF- autocorrelation plots of the monthly maximum temperatures of Sylhet after differencing are shown in Figure 2 and give an indication of the significant orders $p$ and $q$ to be used in model-setup.

Once the most appropriate order of the ARMA- or ARIMA model is specified, the AR- and MA-coefficients in Eq. (1) are estimated in step ii) of the Box-Jenkins process by linear least-squares. The goodness of the best models is evaluated using the AIC (Akaike Information Criterion).

Table 1 lists the ARIMA models obtained for the maximum and minimum temperatures time series at Sylhet and Sreemongal stations, together with their corresponding AIC-values.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Station Name</th>
<th>ARIMA Model</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max. Temp.</td>
<td>Sylhet</td>
<td>(1, 1, 1) (1, 1,1)</td>
<td>1211.30</td>
</tr>
<tr>
<td>Min. Temp.</td>
<td>Sylhet</td>
<td>(1, 1, 1) (0, 1,1)</td>
<td>1810.38</td>
</tr>
<tr>
<td>Max. Temp.</td>
<td>Moulvibazar</td>
<td>(1, 1, 0) (1, 1,1)</td>
<td>1836.66</td>
</tr>
<tr>
<td>Min. Temp.</td>
<td>Moulvibazar</td>
<td>(0, 1, 1) (1, 1,1)</td>
<td>914.27</td>
</tr>
</tbody>
</table>

As prescribed by step iii) Box-Jenkins methodology, the identified ARIMA model must be diagnostically checked for its appropriateness, by looking at the ACF and PACF of the model residuals. These are shown for the monthly maximum temperatures at Sylhet station in Figure 3. As the spikes at the different lags in the ACF and PACF plots in the figure are within the statistical confidence bands, the ARIMA (1,1,1) (1,1,1) model for this time series is adequate. Similarly good results have been obtained for the other time series analyzed.

The final step (iv) and ultimate goal of ARIMA modeling is then the forecasting of the time series for one or more future time steps ahead [9], using Eq. 1, with the coefficients determined in step (ii) and values of the time series from previous months.

Figures 4 and 5 show one-month-ahead predictions for the monthly maximum temperatures for years 2010 and 2011 at Sylhet and Moulvibazar station, respectively. From the visual inspection of these

Figure 2: ACF (top panel) and PACF (bottom panel) of monthly maximum temperatures at Sylhet station after seasonal differencing.

Figure 3: ACF (top panel) and PACF (bottom panel) of residuals of monthly maximum temperatures at Sylhet station.
graphs one may note that the predicted time series is very close to the other three time series investigated which indicates that the ARIMA-model is a valuable tool for short-term forecasting of the two meteorological variables max. and min. temperature.

![Figure 4: Observed and ARIMA-predicted monthly maximum temperatures at Sylhet station.](image)

**Figure 4:** Observed and ARIMA-predicted monthly maximum temperatures at Sylhet station.

![Figure 5: Observed and ARIMA-predicted monthly maximum temperatures at Moulvibazar station.](image)

**Figure 5:** Observed and ARIMA-predicted monthly maximum temperatures at Moulvibazar station.

**CONCLUSIONS**

The temperature time series fitted by the ARIMA model for the two selected stations can be used for estimating missing temperature values and for forecasting. Thus, the Box-Jenkins methodology can help decision makers to establish better strategies and to set up priorities for arming themselves against upcoming weather changes which may have effects last, but not to the least, on the water resources in the Sylhet division.

**REFERENCES**


