

SARIMA modeling of the monthly average maximum and minimum temperatures in the eastern plateau region of India

ENAKSHI SAHA, ARNAB HAZRA and PABITRA BANIK

Agricultural and Ecological Research Unit, Indian Statistical Institute, Kolkata – 700 108, India

(Received 5 June 2015, Accepted 10 December 2015)

e mail : enu2.saha@gmail.com

सार – वर्ष 1990-2011 के लिए भारत के गिरिडीह में एकत्रित किए गए मासिक औसत अधिकतम और न्यूनतम तापमान आंकड़ा सेट में SARIMA समय श्रृंखला मॉडल फिट किया गया। समय श्रृंखला प्लॉटों से हमने देखा कि दोनों श्रृंखलाओं के पैटर्न काफी अलग-अलग हैं : अधिकतम तापमान श्रृंखलाओं में लगभग सभी वर्षों में तीव्र चरम सीमा रही है जबकि न्यूनतम तापमान की श्रृंखलाओं में यह सही नहीं है अतः दोनों श्रृंखला के मॉडलों को अलग-अलग तैयार किया गया है। मासिक श्रृंखलाओं के आंशिक स्वतः सहसंबंध (PACF) और प्रेक्षित स्वतः सहसंबंध कार्य (ACF) के आधार पर SARIMA मॉडलों का चयन किया गया है। तीन जांचों (अर्थात मानक त्रुटि, अवशिष्टों के ACF और PACF तथा Akaike सूचना मापदंड (AIC) बायेसियन सूचना मापदंड (BIC) और संशोधित Akaike सूचना मापदंड (AICc) के सहयोग से अधिकतम संभावित पद्धति का उपयोग करते हुए मॉडल के प्राचल प्राप्त किए गए। मानकीकृत अवशिष्टों, अवशिष्टों के ACF, माननीकृत अवशिष्टों के Q-Q प्लॉट और Liung-box आंकड़ों के p - मानों के साथ नैदानिक जाँच करते हुए चयनित मॉडलों की पर्याप्तता निर्धारित की गई। अन्ततः भारत के पूर्वी पठार के क्षेत्रों के लिए मासिक औसत अधिकतम और न्यूनतम तापमानों का पूर्वानुमान लगाने के लिए क्रमशः ARIMA (1; 0; 2) × (0; 1; 1)₁₂ और ARIMA (0; 1; 1) × (1; 1; 1)₁₂ मॉडलों को चुना गया।

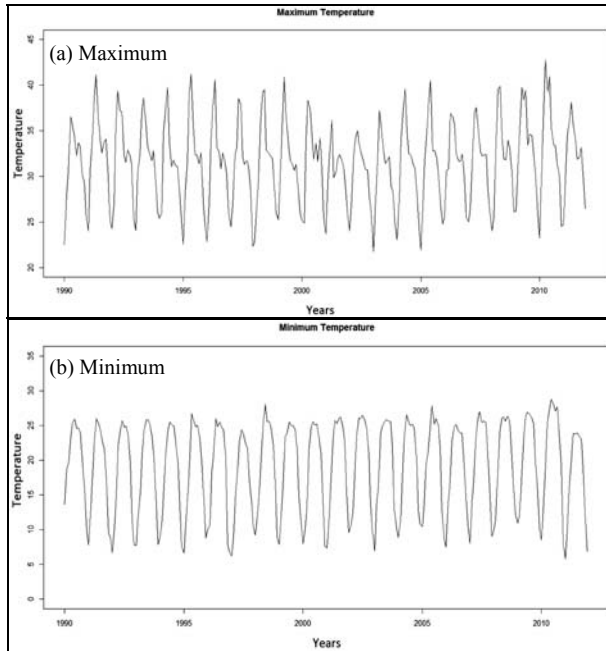
ABSTRACT. The SARIMA time series model is fitted to the monthly average maximum and minimum temperature data sets collected at Giridih, India for the years 1990-2011. From the time-series plots, we observe that the patterns of both the series are quite different; maximum temperature series contain sharp peaks in almost all the years while it is not true for the minimum temperature series and hence both the series are modeled separately (also for the sake of simplicity). SARIMA models are selected based on observing autocorrelation function (ACF) and partial autocorrelation function (PACF) of the monthly temperature series. The model parameters are obtained by using maximum likelihood method with the help of three tests [*i.e.*, standard error, ACF and PACF of residuals and Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC) and corrected Akaike Information Criteria (AICc)]. Adequacy of the selected models is determined using diagnostic checking with the standardized residuals, ACF of residuals, normal Q-Q plot of the standardized residuals and p -values of the Ljung-Box statistic. The models ARIMA (1; 0; 2) × (0; 1; 1)₁₂ and ARIMA (0; 1; 1) × (1; 1; 1)₁₂ are finally selected for forecasting of monthly average maximum and minimum temperature values respectively for the eastern plateau region of India.

Key words – Seasonal ARIMA model, Autocorrelation, Partial autocorrelation, Normal Q-Q plot, Ljung-Box statistic.

1. Introduction

The air temperature is one of the most important meteorological factors from the environmental, ecological and agricultural point of view. Modeling the variability of surface air temperature and producing reliable forecasts underlie the foundation of sound agricultural policies, particularly important for the eastern plateau region of India because the income of a major portion of the local people is low and completely dependent on agriculture. Moreover, temperature is a critical input parameter in many eco-environmental models in the fields of crop growth simulation (Verdoodt *et al.*, 2004; Bechini *et al.*, 2006), agro-ecological zoning (Caldiz *et al.*, 2001; Ye

et al., 2008) and food security assessment (Ye and Van Ranst, 2009; Ye *et al.*, 2012). Policy analysis using these ecosystem models is only possible with an accurate prediction of future temperatures. Several efforts have been made in the statistical time series modeling of temperature variations using monthly average temperature records (Hansen *et al.*, 2006; Rahmstorf *et al.*, 2007). Among them, the univariate time-series models have gained relative popularity in recent years, partly due to the complexity of mainstream climate models, which are strongly constrained by the current knowledge of the physical climate system (IPCC, 2003). One subcategory of the univariate models, namely the structural time-series models (Lee and Sohn, 2007), has become quite popular



Figs. 1(a&b). Monthly average maximum and minimum temperature series in Giridih, Jharkhand from 1990 to 2012

due to its trend-detecting capability. In general, a structural time-series model comprises a deterministic trend plus random residuals about the trend, where the residuals are assumed to represent natural variability and can be viewed as a realization of an autoregressive integrated moving average (ARIMA) process (Romilly, 2005) and seasonal ARIMA (SARIMA) is also important in case of meteorological parameters (Bender and Simonovle, 1994; Montanari *et al.*, 2000; Trawinski and Mackay, 2008; Khajavi *et al.*, 2012; Meshram *et al.*, 2012).

We fit a SARIMA, *i.e.*, seasonal (S) auto-regressive (AR) integrated (I) moving average (MA) time series model to the monthly average maximum and minimum temperature datasets collected at Giridih, India for the years 1990-2011. From the time-series plots [Figs. 1(a&b)], we observe that the patterns of both the series are quite different; maximum temperature series contain sharp peaks in almost all the years while it is not true for the minimum temperature series and hence both the series are modelled separately. SARIMA models were selected based on observing ACF and PACF of the monthly temperature series. The model parameters were obtained by using maximum likelihood method, and the variable selection has been done based on AIC, BIC and AICc. Diagnostic checking has been done with the standardized residuals, ACF of residuals, normal Q-Q plot of the standardized residuals and p-values of the Ljung-Box statistic. The ARIMA model that passed the adequacy

test was selected for forecasting. The data for the 22 years, 1990-2011, have been used for model fitting and data of the year 2012 have been used for model verification.

The paper is organized as follows. In the Data and methodology section, we describe the data set, the model that we fit to the data and the procedure of diagnostic checking. In the Results and Discussion section, SARIMA models are fitted to the monthly maximum and the minimum temperature series, tabulate the estimates of the model parameters with AIC, BIC and AICc values and check the model diagnostics. In the Conclusion section, we include some concluding remarks.

2. Data and methodology

The data : Daily temperature data of Giridih (24° 18' N, 86° 30' E) for 23 year period 1990 - 2012 are used in this work, which have been collected by Indian Statistical Institute, Giridih and the monthly data sets used in this paper are prepared based on the average monthly maximum and minimum temperatures. The temperature data sets are collected using instruments conforming to the specified criteria of India Meteorological Department.

The model identification : We assume that the data is stationary over the years, *i.e.*, there is no yearly trend and the only variation arises due to seasonality. As we are dealing with only 22 years of data, this assumption is quite reasonable and [Figs. 1(a&b)] is also an evidence for it. First step to identify the model is studying the ACF and PACF plots. Once a set of possible models is identified, the best model is to be identified based on AIC, BIC and AICc values. To be sure about the identification of the correct model, diagnostic checking with the standardized residuals, ACF of residuals, normal Q-Q plot of the standardized residuals and p-values of the Ljung-Box statistic is required. We discuss these steps sequentially in the following. For more details, please refer to Shumway and Stoffer (2011).

Steps 1

Identification of the possible set of models based on ACF and PACF

The sample autocorrelation function (ACF) of a stationary process is given by

$$\hat{\rho}(h) = \frac{\sum_{t=1}^{T-h} (Y_{t+h} - \bar{Y})(Y_t - \bar{Y})}{\sum_{t=1}^T (Y_t - \bar{Y})^2}$$

where, T denotes the total number of months at which we have observations; *e.g.*, if we have monthly data

for 22 years then $T = 22 \times 12 = 264$, Y_t denotes the observation for the t -th month; \bar{Y} denotes the mean of the observations, h denotes the lag size.

The sample partial autocorrelation function (PACF) of a stationary process is given by :

$$\hat{\phi}_{11} = \hat{\rho}(1)$$

$$\hat{\phi}_{hh} = \frac{\sum_{t=1}^{T-h} (Y_{t+h} - \hat{Y}_{t+h})(Y_t - \hat{Y}_t)}{\sum_{t=1}^T (Y_t - \hat{Y}_t)^2}$$

Both $(Y_{t+h} - \hat{Y}_{t+h})$ and $(Y_t - \hat{Y}_t)$ are uncorrelated with $(Y_{t+h-1}, \dots, Y_{t+1})$, *i.e.*, \hat{Y}_{t+h} and \hat{Y}_t are the estimators of Y_{t+h} and Y_t after fitting two linear regressions with $Y_{t+h-1}, \dots, Y_{t+1}$ as the dependent variables.

The large sample standard deviation of both the quantities, ACF and PACF, are equal to $\frac{1}{\sqrt{T}}$ with mean zero under the general white noise assumption, where, T is as defined earlier, the total number of months we considered. Thus, under the large sample Gaussian assumption in both the cases, the quantities are assumed to be significant at level 0.05 (standard statistical assumption) if

$$|\hat{\rho}(h)| > \frac{1.96}{\sqrt{T}}; |\hat{\phi}_{hh}| > \frac{1.96}{\sqrt{T}}$$

Our target is to make the time series stationary and thus, if the above quantities appear to be significant (which is quite natural if seasonal component is not removed), we need to consider the seasonal differences using Backshift operators.

If we observe that $\hat{\rho}(h)$ are significant for $h = 1, 2, \dots, q$ and becomes insignificant afterwards while $\hat{\phi}_{hh}$ tails off over $h = 1, 2, \dots$, but does not become insignificant after some certain value of h , then we identify the moving average (MA) order of the model to be q . On the other hand, if we observe that $\hat{\phi}_{hh}$ are significant for $h = 1, 2, \dots, p$ and becomes insignificant afterwards while $\hat{\rho}(h)$ tails off over $h = 1, 2, \dots$, but does not become insignificant after some certain value of h , then we identify the autoregressive (AR) order of the model to be p . In case, both $\hat{\rho}(h)$ and $\hat{\phi}_{hh}$ tails off

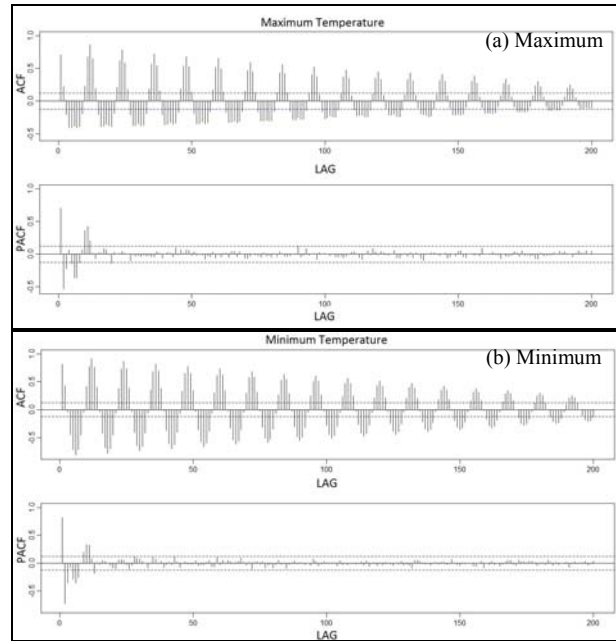


Fig. 2(a&b). ACF and PACF plots for the monthly average maximum and minimum temperature series

over $h = 1, 2, \dots$, but do not become insignificant after certain values of h , then the model would be autoregressive and moving average together. The orders need to be determined based on observing the ACF and PACF plots and checking the AIC, BIC and AICc which is almost always needs to check for comparison and described in the next step. If it appears that there are q spikes in the ACF plot and p spikes in the PACF plot, the ARMA model with autoregressive order p and moving average order q is reasonable to fit the model well and should be checked. If the seasonal components are present, the same procedure needs to perform with multiplying h by s where s denotes the seasonal frequency and for a monthly data, $s = 12$ and the models would be SAR, SMA and SARMA in place of AR, MA and ARMA respectively. Often we need to consider the seasonal or monthly differences of the time series, for example, we may need to work with the differences time series, *e.g.*, $Z_t = Y_t - Y_{t-1}$ or $Z_t = Y_t - Y_{t-12}$. Then, the models are called integrated (and hence, I) models and the most generalized case SARMA would become a SARIMA model.

Step 2

Identification of the most preferable model based on AIC, BIC and AICc values

In case we increase the number of parameters to make the model more flexible, the models with more

parameters, perform better due to flexibility and return a smaller estimate of the variance component. Thus, for a variable selection problem like ours, the idea is to judge based on some information criteria, for example AIC, BIC and AICc given by :

$$AIC = \log(\hat{\sigma}^2) + \frac{T+2k}{T}; \quad BIC = \log(\hat{\sigma}^2) + \frac{k \log(T)}{T};$$

$$AICc = \log(\hat{\sigma}^2) + \frac{T+k}{T-k-2}$$

where, k denotes the total number of parameters in the model and $\hat{\sigma}^2$ is the estimated variance.

The model with minimum information criteria values are considered to be the better ones. We need to make our choice set smaller from the set obtained in Step 1.

Step 3

Diagnostic checking of the models chosen based on information criteria

Once the preferred model is identified, Standardized residuals are looked at. According to our model assumption, observations are normally distributed and thus, the standardized residuals should be standard normally distributed. Thus, one should look into the normal Q-Q plot of the standardized residuals. Now, if a model does not fit the data very well, the errors will no more remain uncorrelated and like a time series depends on its previous values, the errors will remain uncorrelated as well (*e.g.*, the monthly average temperature of February, 2012 is expected to be dependent on a few earlier months, say, January, 2012 and December, 2011 and if we could not fit a proper model, the error corresponding to February, 2012 will remain dependent on the errors corresponding to the months January, 2012 and December, 2011). One should also inspect the sample autocorrelations of the residuals, say, at $\hat{\rho}_e(h)$ lag h , because the values should be very close to zero by white noise assumption. It is possible that the individual $\hat{\rho}_e(h)$ values are negligible over a set of values of h but they are not negligible altogether. Thus, we should look into an aggregated measure that considers $\hat{\rho}_e(h)$ values over a set of values of h , say, $h = 1, 2, \dots, H$. One might suspect that the model does not fit the data very well and thus, the errors have high autocorrelation over a sufficiently large number of lags and thus, the maximum lag H is to be chosen large enough; typically H is set as 20.

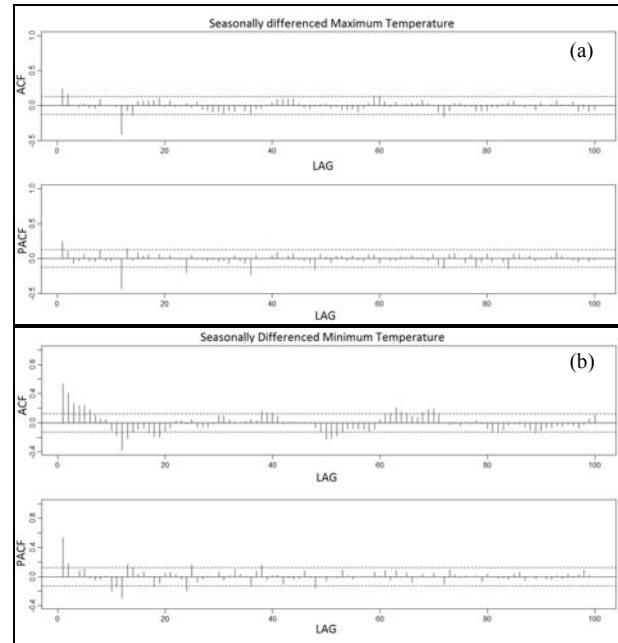


Fig. 3(a&b). ACF and PACF plots for the difference monthly average maximum and minimum temperature series

To check that scenario, the Ljung-Box-Pierce Q-statistic given by :

$$Q = T(T+2) \sum_{h=1}^H \frac{\hat{\rho}_e^2(h)}{T-h}$$

But checking over several values of H is more justified; *e.g.*, $H = 5, 6, \dots, 35$ and a software like R returns the values of Q over a sufficiently large number of choices of H . Under the null hypothesis of model adequacy, asymptotically, $Q \sim \chi_{H-p-q}^2$, where, p is the number of model parameters for the AR component and q is the number of model parameters for the MA component of the SARIMA model. Thus, we would reject the null hypothesis at level α if the value of Q exceeds the $(1 - \alpha)$ -th quantile of the χ_{H-p-q}^2 distribution. All model parameters are to be estimated using the maximum likelihood estimation procedure.

3. Results and discussions

In this section, the described method is implemented to our dataset which also illustrates the procedure described. Within each step, the results for the maximum and minimum temperature are described separately. In Figs. 2(a&b), the ACF and PACF plots are provided for the average maximum and the average minimum temperature series. Mainly

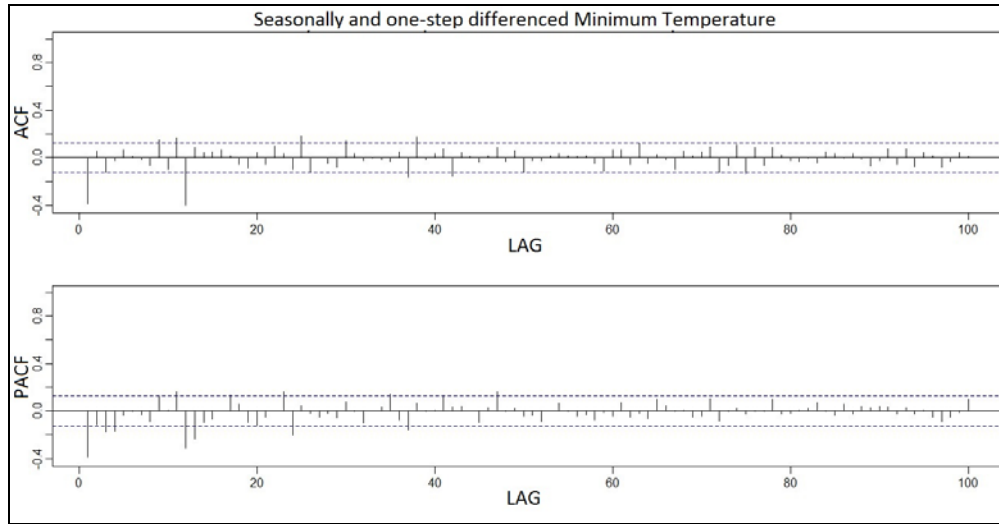


Fig. 4. ACF and PACF plots for the X_t series derived from the series of monthly average minimum temperature

observable component is the seasonal component with frequency = 12.

In the Step 1, our aim is to identify the seasonal, autoregressive and moving average orders of the SARIMA model. As the monthly average temperature series is very likely vary seasonally and, for example, the average temperature for January, 2012 is likely to be similar to the average temperature of January, 2011 and similarly for other months, following the Step 1, first the seasonal component is removed for making it stationary. The transformed variables are considered, *i.e.*, $Z_t = Y_t - Y_{t-12}$ and consider the ACF and PACF plots of Z_t ; $t = 13, \dots, T$ in Fig. 3.

For the monthly average maximum temperature series, first, concentrating on the seasonal ($s = 12$) lags, the characteristics of the ACF and PACF of these time series tend to show a strong peak at $h = 1s$ in the autocorrelation function, combined with peaks at $h = 1s$; $2s$; $3s$; $4s$ in the partial autocorrelation function. It appears that either.

- (i) the ACF is cutting off after lag $1s$ and the PACF is tailing off in the seasonal lags, or
- (ii) the ACF and PACF are both tailing off in the seasonal lags.

Following the Step 1, these suggest that either (i) an SMA of order $Q = 1$, or (ii) an SARMA of orders $P = 3$ (because of the three spikes in the PACF) and $Q = 1$.

Next, inspecting the ACF and the PACF at the within season lags, $h = 1, \dots, 11$, it appears that the ACF cuts off at lag 2 and the PACF cuts off at lag 1. This result indicates that we should consider fitting a model with $p = 1$ and $q = 2$ for the non-seasonal components.

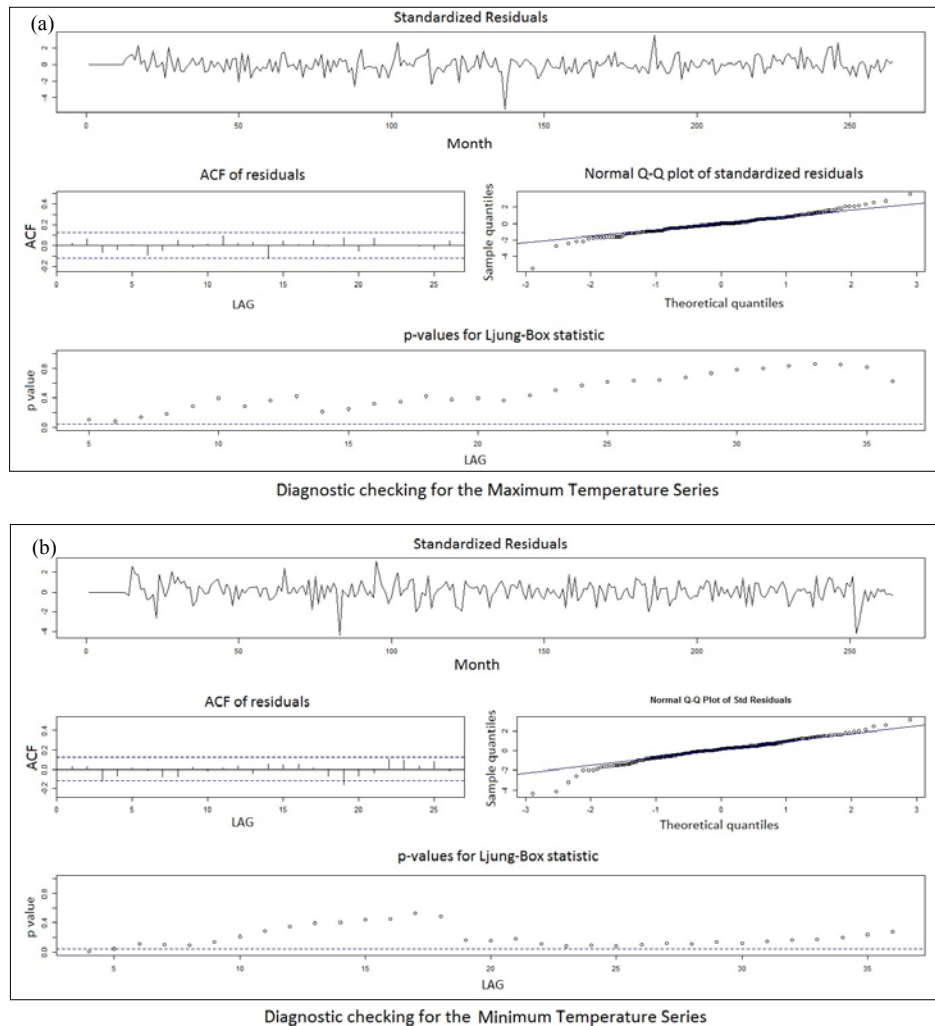
Fitting the three models suggested by these observations, we obtain:

- (i) ARIMA (1; 0; 2) \times (0; 1; 1)₁₂: AIC = 1.868; AICc = 1.877; BIC = 0.936
- (ii) ARIMA (1; 0; 2) \times (3; 1; 1)₁₂: AIC = 1.909; AICc = 1.919; BIC = 1.017

Following Step 2, The ARIMA (1; 0; 2) X (0; 1; 1)₁₂ is the preferred model and the parameter estimates of different components (standard errors of the model parameters are reported within brackets) in this case are given by :

AR1	MA1	MA2	SMA1	Constant	Variance
0.9705	-0.6769	-0.1901	-1.0000	0.0033	2.294
(0.0307)	(0.0664)	(0.0597)	(0.0546)	(0.0043)	-

For the monthly average minimum temperature series, in Figs. 3(a&b), we still observe that the ACF tails off very slowly in between the seasonal lags indicating that taking a first difference is necessary. After taking first difference, the ACF and PACF plots are as follows. We consider the transformed variables $X_t = Z_t - Z_{t-1}$ and consider the ACF and PACF plots of X_t in Fig. 4.



Figs. 5 (a&b). Diagnostic checking for the models finally selected in Step 2 using ACF of residuals, Normal Q-Q plot of the standardized residuals and the p -values of the Ljung-Box test statistic

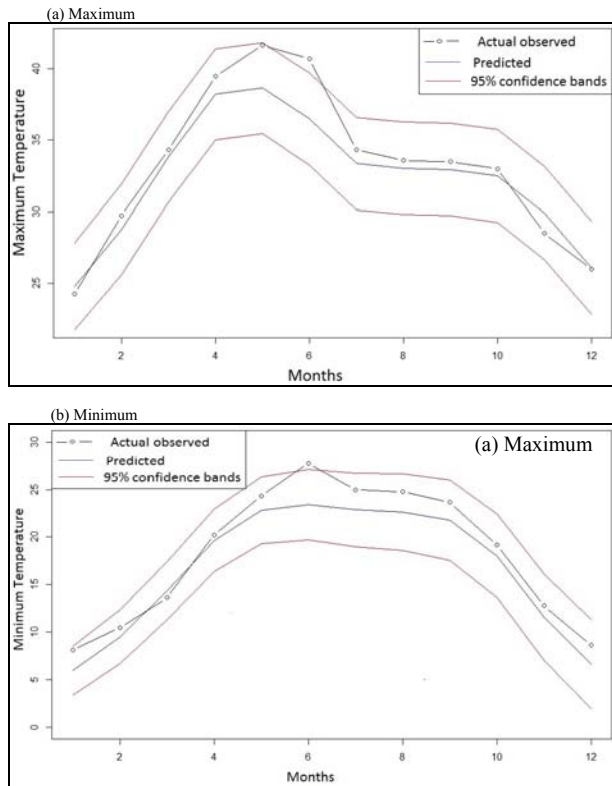
First, concentrating on the seasonal ($s = 12$) lags, the characteristics of the ACF and PACF of this series tend to show a strong peak at $h = 1s$ in the autocorrelation function, combined with peaks at $h = 1s; 2s; 3s; 4s$ (peaks at $h = 1s; 2s$ are more significant and peaks at $h=3s; 4s$ are negligible) in the partial autocorrelation function. It appears that either.

- (i) the ACF is cutting off after lag $1s$ and the PACF is tailing off in the seasonal lags, or
- (ii) the ACF and PACF are both tailing off in the seasonal lags.

This suggests either (i) a SIMA of order $Q = 1$, or (ii.a) an SARIMA of orders $P = 2$ (because of the

two significant spikes in the PACF) and $Q = 1$ (because the first spike is more significant) or (ii.b) an SARIMA of orders $P = 2$ (because of the two significant spikes in the PACF) and $Q = 2$ (considering the second spike also).

Next, inspecting the ACF and the PACF at the within season lags, $h = 1, \dots, 11$, it appears that the ACF cuts off at lag 1 and PACF tails off, or (b) both the ACF and PACF are tailing off. This result indicates that we should either consider fitting a model (a) with both $p > 0$ and $q > 0$ for the non-seasonal components, say $p = 1; q = 1$ (because of significant spikes), or (b) $p = 0; q = 1$. Fitting the six models suggested by these observations, we obtain:



Figs. 6 (a&b). Performance of the selected SARIMA models to the maximum and minimum temperature series. The blue line indicates the predicted values and the red lines indicate the 95% prediction intervals while black lines with bubbles indicate the actually observed values for the year 2012

- (i) ARIMA (0; 1; 1) × (1; 1; 1)₁₂ : AIC = 1.525021; AICc = 1.533182; BIC = 0.5656567
- (ii) ARIMA (0; 1; 1) × (0; 1; 1)₁₂ : AIC=1.554324; AICc = 1.56225; BIC = 0.5814147
- (iii) ARIMA (2; 1; 1) × (1; 1; 1)₁₂ : AIC=1.500685; AICc =1.509498; BIC=0.5684108
- (iv) ARIMA (2; 1; 1) × (0; 1; 1)₁₂ : AIC=1.542679; AICc = 1.551136; BIC = 0.5968603
- (v) ARIMA (2; 1; 2) × (1; 1; 1)₁₂ : AIC = 1.507961; AICc = 1.517194; BIC = 0.5892323
- (vi) ARIMA (2; 1; 2) × (0; 1; 1)₁₂ : AIC = 1.548; AICc = 1.556814; BIC = 0.6157263

The ARIMA (0; 1; 1) × (1; 1; 1)₁₂ is the preferred model and again the parameter estimates of different components (standard errors of the model parameters are reported within brackets) in this case are given by :

MA1	SAR1	SMA1	Variance
-0.5344 (0.0641)	0.0762 (0.0788)	-0.9841 (0.3045)	1.653 -

The diagnostics for the fits in both the cases are displayed in Figs. 5(a&b). We note that all the ACF values of the residuals are non-significant in case of maximum temperature and it is significant for the lag 19 in case of minimum temperature but the peak crosses the confidence interval by a very negligible amount. Normal Q-Q plot signifies that the normality of the residuals is a reliable assumption. Significant deviation is observable at the left tail, but the fitting is perfect for the rest of the support. In both the cases, the *p*-values for the Ljung-Box test statistics are greater than 0.05 for almost all the values of *H*. For *H* = 4, 5, the *p*-values are smaller than 0.05 in case of minimum temperature but that also by a negligible amount. Hence, from the Step 3, we conclude that the finally chosen models from Step 2 qualifies the criteria of Step 3.

Finally, to demonstrate the performance of our model, forecasts based on the fitted models for the next 12 months are shown in Figs. 6(a&b). Except for the observations corresponding to June, all others lie within the 95% prediction intervals which indicates better performance of the chosen SARIMA models. In June 2012, we notice that both the observations, *i.e.*, the observed average maximum temperature and the observed average minimum temperature lie above the upper bound of the prediction intervals which also indicates the highly unpredicted extremely hot summer.

4. Conclusions

Our study indicates that the seasonal ARIMA model is a perfect tool for monthly temperature forecasting. For forecasting monthly average maximum temperature, ARIMA (1; 0; 2) × (0; 1; 1)₁₂ is chosen to be the best performing model while ARIMA (0; 1; 1) × (1; 1; 1)₁₂ performs the best in case of monthly average minimum temperature forecasting. Out of 12 months, the observed values lie outside the prediction interval only for one month corresponding to the recent problem of extremely unpredicted hot summer months in India. The extreme weather prediction is a very different approach from that of ours. Extreme value analysis would be an approach to predict such extremal events.

References

Bechini, L., Bocchi, S., Maggiore, T. and Confalonieri, R., 2006, "Parameterization of a crop growth and development simulation model at sub model component level : An example for winter wheat (*Triticum aestivum* L.)", *Environmental modelling & software*, **21**, 1042-1054.

- Bender, M. and Simonovle, S., 1994, "Time-series modeling for long-range stream flow forecasting", *Journal of Water Resources Planning and Management*, ASCE, **120**, 6, 857-870.
- Caldiz, D. O., Fernandez, L. V. and Struik, P., 2001, "Physiological age index: A new, simple and reliable index to assess the physiological age of seed potato tubers based on haulm killing date and length of the incubation period", *Field Crops Research*, **69**, 69-79.
- Hansen, J., Sato, M., Ruedy, R., Lo, K., Lea, D. W. and Medina-Elizade, M., 2006, "Global temperature change", *Proceedings of the National Academy of Sciences*, **103**, 14288-93.
- IPCC, 2003, "Intergovernmental Panel on Climate Change : Third Assessment Report - Climate Change 2001", World Meteorological Organization, Geneva.
- Khajavi, E., Behzadi, J., Nezami, M. T., Ghodrati, A. and Dadashi, M. A., 2012, "Modeling ARIMA of air temperature of the southern Caspian Sea coasts", *International Research Journal of Applied and Basic Sciences*, **3**, 6, 1279-1287.
- Lee, J. H. and Sohn, K. T., 2007, "Prediction of monthly mean surface air temperature in a region of China", *Advances in Atmospheric Sciences*, **24**, 503-508.
- Meshram, D. T., Gorantiwar, S. D. and Lohakare, A. S., 2012, "ARIMA model for forecasting of evaporation of Solapur station of Maharashtra, India", *Mausam*, **63**, 4, 573-580.
- Montanari, A., Rosso, R. and Taquu, M. S., 2000, "A seasonal fractional ARIMA model applied to the Nile river monthly flows at Aswan", *Water Resources Research*, **36**, 5, 1249-1259.
- Rahmstorf, S., Cazenave, A., Church, J. A., Hansen, J. E., Keeling, R. F., Parker, D. E. and Somerville, R. C. J., 2007, "Recent climate observations compared to projections", *Science*, **316**, DOI: 10.1126/science.1136843.
- Romilly, P., 2005, "Time series modelling of global mean temperature for managerial decision-making", *Journal of Environmental Management*, **76**, 61-70.
- Shumway, R. H. and Stoffer, D. S., 2011, "Time Series Analysis and Its Applications", Springer-Verlag, Third Edition.
- Trawinski, P. R. and Mackay, D. S., 2008, "Meteorologically conditioned time series predictions of West Nile virus vector mosquitoes", *Vector-Borne and Zoonotic Diseases*, **8**, 4, 505-522.
- Verdoodt, A., van Ranst, E. and Ye, L. M., 2004, "Daily simulation of potential dry matter production of annual field crops in tropical environments", *Agronomy Journal*, **100**, 1739-1753.
- Ye, L., Tang, H., Zhu, J., Verdoodt, A. and Van Ranst, E., 2008, "Spatial patterns and effects of soil organic carbon on grain productivity assessment in China", *Soil Use and Management*, **24**, 80-91.
- Ye, L., and Van Ranst, E., 2009, "Production scenarios and the effect of soil degradation on long-term food security in China", *Global Environmental Change*, **19**, 464-481.
- Ye, L., Xiong, W., Li, Z., Yang, P., Wu, W., Yang, G., Fu, Y., Zou, J., Chen, Z., Van Ranst, E. and Tang, H., 2012, "Climate change impact on China food security in 2050", *Agronomy for Sustainable Development*, DOI: 10.1007/s13593-012-0102-0.
-