Mysore J. Agric. Sci., 52 (4) : 784-789 (2018)

Fitting of Probability Distributions for Bias Minimized Modeled Temperature Data

U. SRUTHI, G. B. MALLIKARJUNA AND K. N. KRISHNAMURTHY
Department of Agricultural Statistics, Applied Mathematics and Computer Science
College of Agriculture, UAS, GKVK, Bengaluru - 560 065
E-mail : kkmurthy13@gmail.com

ABSTRACT

The study aims at reducing the bias in the temperature data obtained from the satellite models by comparing with that obtained from the observatories. Data obtained from the observatories are always more accurate than those from the satellite model. Bias correction methods such as difference method (DM) and modified difference method (MDM) were attempted to minimize bias of the satellite modeled temperature data compared to the data recorded at observatories. Best bias correction method was identified based on the coefficient of variation. MDM recorded lower CV in the corrected modeled data based on daily and weekly maximum and minimum temperatures. This indicated its use in smoothening of the data. Difference Method (DM) recorded lower CV in the corrected modeled data of weekly minimum temperature and monthly data of maximum and minimum temperatures. Generally, studies of temperatures are more related to months rather than week/years. For the bias corrected modeled (satellite data) monthly data, distributions functions were fitted and their goodness of fit was verified using Chi-square test. No generalized single model was found to be the best fit for both the climatic parameters. Cauchy, Gamma and chi-squared distributions were resulted as the best fit more than once in the analysis for Maximum temperature. For the Minimum temperature, Log-logistic distribution was more in number compared to other distributions.

Keywords : Probability distribution, Modified difference method, CV

GLOBAL climate models (GCMs) are basic tools for predicting future climate to enable a better understanding of climate change. The role of statistical methodology for predicting the weather parameters is considered to be most important for their precise estimates. Although, high-speed computers, meteorological satellites and weather radars are tools that had played major roles in improving weather forecasts. But the Improvement in initial conditions is the result of an increased number of observations and better use of the observations in computational techniques.

Predicted climatic parameters will have a significant impact on water resources and hydrology. Any study related to this requires temporal and spatial data on climatic parameters. Monitoring and understanding temporal and spatial data of climatic parameters can assist in better preparation for drought conditions. Ground stations (observatories) are too sparse to achieve the coverage needed for accurate analysis of climatic parameters, especially as spatial variability. Climatic parameters monitoring at ground stations over most places does not provide data with the speed, reliability and accuracy required for early warning of droughts. Data collection of climatic parameters in remote area is also limiting factor with the ground stations. To overcome this, satellite data are used for the estimation, but they are not so reliable because of large distance capture. Climate models (based on satellite data) are not perfect in providing simulated climatology. They will differ from observed climatology. The model state will drift towards the model climate as the forecast progresses and this drift will be confounded with the climate evolution that is being predicted. For this reason, near-term climate predictions are usually bias corrected. Broadly, bias includes any type of error that is systematic rather than random. In reality, errors in models and data are often systematic rather than random. The bias may be temporal, spatial, seasonal or even situationdependent. The size of the bias depends on the accuracy as well as the frequency of the observations. In statistics, bias is a property of an estimator which, under- or over estimates some quantity. Because of incomplete understanding of the physics of the climate system, different climate modeling groups around the world represent climate processes in different ways in their models. As a result, there are differences in the projections of future climate. This is therefore, a source of uncertainty in climate projections (known as structural error).

To achieve precision in forecasting, data has to be made free of systematic error. This can be done with the introduction of bias correction factors to the data and then statistical models can be fitted to the bias corrected data and then validated for their appropriateness.

MATERIAL AND METHODS

The present study was carried out in Bengaluru Urban district of the South Indian state of Karnataka. It is surrounded by the Bengaluru Rural district on the East and North, the Ramanagara district on the West and the Krishnagiri district of Tamil Nadu on the South. Bengaluru Urban district came into being in 1986, with the partition of the erstwhile Bengaluru into Bengaluru Urban and Bengaluru Rural districts. Bengaluru urban district comes under Eastern dry zone of the 10 agro climatic zones. This zone consists of an area of 1.808 Mha. The annual rainfall ranges from 679.1 - 888.9 mm. More than 50 per cent of it is received during the *kharif* season. The elevation is 800-900 m above the sea level and the soils are red loamy in major areas, lateritic in the remaining areas.

Present study was based on the secondary data on temperature over a period of 9 years (3240 observations from 2008 to 2016) which was collected from AICRP on Agro Meteorology, University of Agricultural Sciences, GKVK, Bengaluru. The conversion of the daily data (3240 observations) of climatic parameter was done by taking the mean of the 7 days data for weekly and then by taking the average of all the values of the particular week of every month for the 9 years. Monthly converted data of temperature have been done; first by smoothing (bias correction) the satellite (modeled) data followed by fitting appropriate distribution function for it.

Statistical Bias Correction Methods

Following two methods were applied to bring the modeled (satellite) data close to the observed with respect to time trend and magnitude.

1. Difference Method (DM)

In this method, the average daily difference of observed and modeled values (Δx) was taken for each Julian day (365 days) averaged from 9 years data (2008-2016). The (Δx) was considered as daily correction factor, which was added to the modeled uncorrected (satellite) value ($X\ model_{uncor}$) to correct it ($X\ model_{cor}$) so that the value approaches the observed ones.

$$X\ model_{cor} = X\ model_{uncor} + (\Delta x)$$

2. Modified difference method (MDM)

This method was similar to the difference method; However some statistical parameters were added to improve the correction function. For example, in case of rainfall, correction for μ and ó were added which aimed at shifting and scaling to adjust the μ and $ó^{2}$.

$$X\ model_{cor} = (X\ model_{uncor} + \Delta X) \times (\sigma X_{obs} / \sigma X_{mod})$$

The correction capability of the correction functions was tested by using the coefficient of variation (CV).

Coefficient of variation = (sd/mean) *100

Fitting of Probability Distributions

Climatic parameters were highly variable in a given period. Hence, there was a need for both on long term (monthly/yearly) as well as short term (weekly) basis analysis. Generally, studies of temperatures are more related to months rather than week/years, here study was restricted to monthly data of temperature. The distributions *viz*., Gamma, Lognormal, Weibull, Pareto etc. were used and goodness of fit was done with the help of chi square test.

TABLE 1

Coefficient of variation for different data of Maximum Temperature

Periods / Parameters	Daily				Weekly				Monthly			
	Actual	MU	MC (DM)	MC (MDM)	Actual	MU	MC (DM)	MC (MDM)	Actual	MU	MC (DM)	MC (MDM)
Mean	29.8023	29.9692	30.1168	30.1123	29.8011	29.9670	30.0387	30.5965	29.7764	29.9365	30.1432	30.2298
Std	3.1361	6.2494	6.2715	6.2697	2.7792	3.5099	3.5098	3.5667	2.5180	2.7939	2.7939	2.8331
CV(%)	10.5231	20.8527	20.8239	20.8210	9.3258	11.7122	11.6843	11.6571	8.4565	9.3328	9.2688	9.3719

TABLE 2

Coefficient of variation for different data of Minimum Temperature

Periods / Parameters	Daily				Weekly				Monthly			
	Actual	MU	MC (DM)	MC (MDM)	Actual	MU	MC (DM)	MC (MDM)	Actual	MU	MC (DM)	MC (MDM)
Mean	19.6617	19.3898	19.7093	19.9049	19.6607	19.3886	19.6762	19.6798	19.6233	19.3414	19.4964	19.5609
Std	3.9587	5.0910	5.1021	5.1036	2.4258	2.9448	2.9448	2.9612	1.9888	2.3587	2.3587	2.4141
CV(%)	20.1342	26.2561	25.8865	25.6397	12.3384	15.1884	14.9663	15.0466	10.1346	12.3949	12.0980	12.3413

Actual: Data from observatories MU: Model Uncorrected MC: Model Corrected

DM: Difference Method MDM: Modified Difference

Method

Results and Discussion

For the smaller area, the data obtained from the observatories were always more accurate than that of the satellite models. It has got many limitations mainly in the coverage of area and timely collection. Alternatively, remote sensing data were collected for a larger area because of its mechanism of sending radiations from the far away distance. In view of the distance capture, we can observe some bias (error) in the data obtained by the satellite. To reduce these bias different corrective methods were adopted to identify the suitable corrective method for each data set separately. The corrected data set with minimum CV value was recognized as the best correction method. Coefficient of variation for all the data pertains to Maximum and Minimum temperature are provided in Tables 1 and 2.

Bias corrected results for maximum temperature are presented in Table 1. Modified Difference Method for the daily (20.8210), for weekly (11.6571) and Difference Method for Monthly (9.2688) have recorded least CV. For the minimum temperature, bias

Table 3
Empirical distribution and best fitted Probability distributions of monthly Maximum Temperature

Month	Mean	Std deviation	Max	Min	Range	Best fitted Probability distribution with x^2 value	Parameter values
January	28.6944	0.6187	29.9492	2.4329	27.5162	Triangular ($x^2 = 0.4753^{NS}$)	m = 28.479 a = 27.370 b = 30.258
February	31.2291	1.0098	33.1370	3.4832	29.6537	Log Logistic ($x^2 = 0.0808^{NS}$)	α = 50.614 β = 31.147
March	33.8415	0.9007	35.3057	2.8796	32.4260	Cauchy ($x^2 = 0.0340^{NS}$)	σ = 0.6063 μ = 33.587
April	34.7261	0.9927	35.9125	5.7504	30.1620	Cauchy ($x^2 = 0.3803^{NS}$)	σ = 0.3184 μ = 34.853
May	34.4488	6.6659	50.1410	38.0353	12.1058	Lognormal (3P) ($x^2 = 0.3762^{NS}$)	σ = 1.219 μ = 0.050 γ = 32.048
June	29.8679	1.2784	33.2802	5.1995	28.0807	Gen. Gamma (4P) ($x^2 = 0.0245^{NS}$)	k = 1.918 α = 0.597 β = 2.835 γ = 28.066
July	28.5156	0.5734	30.0287	2.7144	27.3143	Gamma ($x^2 = 0.0325^{NS}$)	α = 24.729 β = 0.011
August	28.3480	0.7355	29.5252	3.7258	25.7994	Chi-squared (2P) ($x^2 = 0.0275^{NS}$)	υ = 3 γ = 25.799
September	28.5213	0.7484	29.9480	3.2999	26.6480	Logistic ($x^2 = 0.1369^{NS}$)	σ = 0.412 μ = 28.521
October	29.0224	0.7454	30.1983	4.0016	26.1966	Cauchy ($x^2 = 0.1833^{NS}$)	σ = 0.282 μ = 29.121
November	27.9001	0.8891	29.1078	4.7694	24.3384	Chi-Squared ($x^2 = 0.0058^{NS}$)	υ = 27
December	27.6249	0.4120	28.3439	1.5009	26.8430	Normal ($x^2 = 0.1330^{NS}$)	m = 0.412 β = 27.625

NS Not significant

TABLE 4

Empirical distribution and best fitted Probability distribution of monthly Minimum Temperature

Month	Mean	Std deviation	Max	Min	Range	Best fitted Probability distribution with x^2 value	Parameter values
January	16.4832	0.6760	2.4915	17.9660	15.4745	Weibull (x^2 = 0.3831^{NS})	α^3 = 28.388 β = 16.735
February	17.7838	1.1108	3.8231	19.8661	16.0430	Weibull (x^2 = 0.1881^{NS})	α = 18.158 β = 18.207
March	20.5416	1.0361	3.9727	22.5486	18.5759	Cauchy (x^2 = 0.0352^{NS})	σ = 0.635 μ = 20.297
April	23.8285	5.2698	29.9466	51.5897	21.6430	Exponential (2P) (x^2 =2.5082 E-11^{NS})	λ = 0.457 γ = 21.643
May	22.6574	0.5929	2.3659	23.8029	21.4370	Triangular (x^2 = 0.1980^{NS})	m = 22.447 a = 21.316 b = 24.111
June	21.1296	0.4881	1.7928	22.2654	20.4726	Log-Logistic (3P) (x^2 = 0.2303^{NS})	α = 2.864 β = 0.765 γ = 20.249
July	20.8218	0.9899	4.4357	23.2485	18.8128	Cauchy (x^2 = 1.0591^{NS})	σ = 0.281 μ = 20.499
August	20.2418	0.6834	3.2947	21.2674	17.9727	Pareto 2 (x^2 = 0.0589^{NS})	α = 189.810 β = 332.590
September	19.9181	1.0592	3.7818	21.0229	17.2411	Log-Logistic (x^2 = 0.9748^{NS})	α = 25.066 β = 19.851
October	20.0945	0.6292	2.9298	21.1196	18.1897	Log-logistic (3P) (x^2 = 0.1553^{NS})	α = 16.501 β = 5.683 γ = -5.683
November	18.6618	0.6583	2.4696	19.8271	17.3575	Triangular (x^2 = 0.5844^{NS})	m = 18.489 a = 17.184 b = 20.138
December	17.0400	1.0193	3.9918	18.9648	14.9729	Cauchy (x^2 = 0.2130^{NS})	σ = 0.670 μ = 17.068

NS Not significant

corrected results are presented in Table 2. Modified Difference Method for the daily (25.6397), Modified Difference Method for weekly (14.9663) and monthly (12.0980) recorded least CV. These values are low as compared to uncorrected model. The results indicated that, the bias correction methods minimizes the variation and thus corrected data can be used for further fitting of the appropriate distribution.

Bias correction of weather parameters was conducted by Li *et al*. (2010) by quantile-based mapping method. Kaur *et.al.* (2015) has also used difference method, modified difference method and statistical bias corrective method to minimize bias in rainfall, temperatures (maximum and minimum). They opined that Schwartz Bayesian criterion (SBC) method for down scaling is more valid as compared to other methods. They have used Root means squared error (RMSE) to evaluate the bias correction methods.

For monthly bias corrected data of maximum and minimum temperatures, probability distribution models were attempted. The empirical distribution and the most appropriate probability distribution for the maximum and minimum temperature value of each month are given in Tables 3 and 4 along with the best fitted distribution which was identified based on the chi-square test statistic value. Using the identified distributions predictions can be made.

Alexander *et al.* (2006) have presented the up-to-date global picture of trend in maximum and minimum temperature and precipitation to study the trend, and probability distributions of indices were derived. For the monthly maximum temperature different probability distributions were fitted and the best among them was recorded based on the goodness of fit test by the chi-square test. Donat and Alexander (2012) have investigated changes in the probability density functions, by considering both maximum and minimum temperatures.

From the current study (analyzed by XLSTAT), it could be inferred that, though the observatories data are more reliable, they have limitation in coverage of large area, timely completion, *etc*. To overcome this, satellite data are used for the estimation, but they are not so reliable because of large distance capture. Therefore, here an attempt was made to identify the correction method to reduce this bias in the data (model uncorrected) obtained from the satellites. Identification was made by comparing CV of the data obtained by the corrected methods. It could be inferred that, bias in the model (satellite) data can be smoothened by employing the correction factor through Difference/ Modified Difference Method. No generalized single model was found as best fit for both the climatic parameters. Cauchy, Gamma and chi-squared distributions were resulted as the best fit more than once in the analysis for Maximum temperature. For the Minimum temperature Log-logistic distribution was more in number compared to other distributions.

REFERENCES

ALEXANDER, L. V., ZHANG, X., PETERSON, T. C., CAESAR, J., GLEASON, B., KLEIN TANK, A. M. G. AND TAGIPOUR, A., 2006, Global observed changes in daily climate extremes of temperature and precipitation. *J. Geophys. Res.,* **3,** D05109 : 1 - 22.

DONAT, M. G. AND ALEXANDER, L. V., 2012, The shifting probability distribution of global daytime and night time temperatures. *Geophys. Res. Lett.*, **39,** L14707 **:** 1 - 4.

KAUR, S., JALOTA, S. K., KAUR, H., VASHISHT, B. B., JALOTA, U. R. AND LUBANA, P. P. S., 2015, Evaluation of statistical corrective methods to minimize bias at different time scale in a regional climate model driven data. *J. Agro Meteor.*, **13** (1) : 29-35.

LI, H., SHEFFIELD, J. AND ERIC, F. WOOD., 2010, Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching. *J. Geophys. Res.*, **115**, D10101: 1 - 20.

(*Received :* December, 2017 *Accepted :* November, 2018)