Mysore J. Agric. Sci., 53 (1) : 48-53 (2019)

Evaluation of Statistical Models for Bias Corrected Rainfall 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 rainfall 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 model rainfall data compared to observatories rainfall data. Best bias correction method is identified based on the coefficient of variation. MDM recorded lower CV compared to DM indicating that MDM was better for all the three periods [daily, weekly and monthly (SMW)] of satellite rainfall data in minimizing the data difference (bias). Probability distributions were attempted for the MDM bias corrected SMW's and monthly rainfall data. Majority of the Standard Meteorological Weekly rainfall data have Gamma and Weibull probability distribution functions which are suitable as identified by the Chi square test. Since, the data has high fluctuations in the monthly bias corrected rainfall values and distributions were positively skewed and Gamma distribution was fitted.

Keywords : Bias rainfall data, Modified difference mentioned & Probability distributions

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. High-speed computers, meteorological satellites and weather radars are tools that had played major roles in improving weather forecasts. However 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 situation-dependent. 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 by using bias correction factors to the data and then statistical models can be fitted to the bias corrected data that will be 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 rainfall 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 available daily data of rainfall was used to compute weekly (SMW) and monthly data. The conversion of the daily data (3240 observations) of climatic parameter is 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. According to Agro meteorology, one year is divided into 52 Standard Meteorological Weeks which are given in Table 1. Collected data of rainfall was subjected to 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 μ and σ of it were added which aimed at shifting and scaling to adjust the μ and σ^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).

$$\text{Coefficient of variation (CV)} = \frac{\sigma}{\bar{X}} \times 100$$

Fitting of Probability Distributions

1. *Rainfall SMW data:* Climatic parameter rainfall is highly variable in a given period. Hence, there was a need for both long term (monthly) as well as short term (weekly) analysis. For the rainfall, the distributions viz., Beta, Gamma, Normal, Lognormal, Weibull, Exponential, Pareto, etc., were used to evaluate the best fit probability distribution. These have been attempted for the SMW data and goodness of fit is done with the help of chi square distribution.

2. *Rainfall Monthly data:* Since most rainfall amounts are small except for a few occasional heavy rains, a typical distribution of the amount of rainfall tends to be positively skewed, and it can be

TABLE 1

Standard Meteorological Weeks
(bold are coming under cropping period)

Week No.	Dates	Week No.	Dates
1	01 Jan - 07 Jan	**27**	**02 Jul - 08 Jul**
2	08 Jan - 14 Jan	**28**	**09 Jul - 15 Jul**
3	15 Jan - 21 Jan	**29**	**16 Jul - 22 Jul**
4	22 Jan - 28 Jan	**30**	**23 Jul - 29 Jul**
5	29 Jan - 04 Feb	**31**	**30 Jul - 05 Aug**
6	05 Feb - 11 Feb	**32**	**06 Aug - 12 Aug**
7	12 Feb - 18 Feb	**33**	**13 Aug - 19 Aug**
8	19 Feb - 25 Feb	**34**	**20 Aug - 26 Aug**
9*	26 Feb - 04 Mar	**35**	**27 Aug - 02 Sep**
10	05 Mar - 11 Mar	**36**	**03 Sep - 09 Sep**
11	12 Mar - 18 Mar	**37**	**10 Sep - 16 Sep**
12	19 Mar - 25 Mar	**38**	**17 Sep - 23 Sep**
13	26 Mar - 01 Apr	**39**	**24 Sep - 30 Sep**
14	02 Apr - 08 Apr	**40**	**01 Oct - 07 Oct**
15	09 Apr - 15 Apr	**41**	**08 Oct - 14 Oct**
16	16 Apr - 22 Apr	**42**	**15 Oct - 21 Oct**
17	23 Apr - 29 Apr	**43**	**22 Oct - 28 Oct**
18	30 Apr - 06 May	44	29 Oct - 04 Nov
19	07 May - 13 May	45	05 Nov - 11 Nov
20	14 May - 20 May	46	12 Nov - 18 Nov
21	21 May - 27 May	47	19 Nov - 25 Nov
22	28 May - 03 Jun	48	26 Nov - 02 Dec
23	**04 Jun - 10 Jun**	49	03 Dec - 09 Dec
24	**11 Jun - 17 Jun**	50	10 Dec - 16 Dec
25	**18 Jun - 24 Jun**	51	17 Dec - 23 Dec
26	**25 Jun - 01 Jul**	52 **	24 Dec - 31 Dec

* Week No. 9 will be 8 days during leap year

** Week No. 52 will always have 8 days

fitted by a gamma distribution with (a< 1). An approximation method suggested by Greenwood and Durand (1960), which is also described by Johnson and Kotz (1970) is attempted.

$$\alpha = \frac{\left(0.5000876 + 0.16488552Y - 0.0544274Y^2\right)}{Y}$$

$$\beta = \frac{\overline{X}}{\alpha}$$

In which $Y = \ln\left(\frac{\overline{X}}{G}\right)$

Where, $\overline{X}$ = the arithmetic mean and

G = geometric mean.

Here α is the shape parameter and β the scale parameter of the distribution

RESULTS AND DISCUSSION

For the smaller area, the data obtained from the observatories are 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. So we can observe some bias (error) in the data obtained by the satellite as view from a distance. To reduce this 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 is recognized as best correction method.

Descriptive statistics for the daily, SMW and monthly for the actual (observatory), model (satellite) and model corrected with DM and MDM are presented in Table 2. Result showed that, CV value is reduced after applying the correction methods, the least CV value was observed to be least for the Modified Difference Method which was 164.143 per cent compared to the rest. For the weekly rainfall data also Modified Difference Method was noticed to be the best corrective method as the CV value was least for it and was found to be 120.892 per cent. Similarly, for the monthly rainfall data, Modified Difference Method has provided the least CV value (81.475) compared to the rest. This indicated that Modified Difference

TABLE 2

Descriptive statistics for daily, weekly and monthly rainfall (mm)

Para meters	Daily				Weekly				Monthly			
	Actual	MU	MC (DM)	MC (MDM)	Actual	MU	MC (DM)	MC (MDM)	Actual	MU	MC (DM)	MC (MDM)
Mean	2.944	3.169	3.494	3.689	2.948	3.169	3.135	3.184	2.929	3.141	3.354	3.456
Std	9.464	6.058	6.058	6.054	4.935	3.845	3.845	3.849	3.008	2.771	2.771	2.816
CV(%)	321.457	191.179	173.389	164.143	167.397	121.357	122.656	120.892	102.723	88.221	82.625	81.475

Method was observed to be the ideal measure for smoothening of the data.

Piani *et al*., (2010) has also applied the bias correction methods by using the Gamma distribution in climatic parameters. This has shown much improvement in the data. They were able to reduce the bias in the data, not only the mean and other moments but also the autocorrelation in the data were improved. Graham *et al.,* (2007) and Weiland et *al.,* (2010) used delta method for the bias correction to correct only the mean of the precipitation. Ines and Hansen (2006) have done the bias correction in rainfall data by using the General Circulation Model (GCM). Vernimmen *et al.,* (2012) have also attempted a simple method to correct the products for bias in real-time to achieve a better agreement with rainfall measured at ground stations.

Different probability distribution functions were evaluated for the Bias corrected data of rainfall. According to Meteorological aspects, the whole year is divided into 52 Standard meteorological weeks. For the cropping period (23rd to 43rd weeks), each of the standard meteorological week for model corrected data, different probability distribution were fitted for the rainfall data. Best probability distribution functions identified using Chi square distributions along with the parameters are presented in Table 3. Table indicated that among the fitted probability distributions for the weekly rainfall, Gamma and Weibull were found to be more suitable for many sowing weeks. It can be inferred that for the weekly rainfall model corrected data gamma distribution fitted well for the 25, 28, 31, 32 and the 41st standard meteorological week and Weibull distribution fitted well for the 26, 29, 33, 36 and the 37th standard meteorological weeks. Lars and Vogel (2008) in their studies have also stated that Gamma (2P) was the best fit distribution for the wet-day daily rainfall based on the common goodness of fit test. Taofik *et al.,* (2013) have found out that Weibull distribution was useful.

For the monthly rainfall data by using long method, the estimates of parameter values of gamma distribution is obtained by using long method and are tabulated in Table 4. The parameter values obtained, indicated that the distributions are shape dominated for all the months except June in which it is scale dominated. The shape dominated data indicate the consistency of the monthly rainfall data. But in the month of June variation of rainfall was high.

Maximum Rainfall was observed in the month of August compared to all other months. When the week wise data is considered, 35th standard Meteorological week was found to have the maximum rainfall. Manikandan *et al.,* (2011) has done similar studies and reported that in Tamil Nadu region, the highest amount of one day maximum rainfall was observed in the month of October. As opined by the author it may be due to the North East rainfall in that state.

From the current study, 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 an attempt was made to identify the correction method to reduce the bias in

TABLE 3

Empirical distribution and best fitted Probability distribution of Rainfall (SMW data) for the cropping period

SMW	Mean	SD	Maximum	Minimum	Range	Suitable Distribution	Parameter Values
W23	71.297	139.364	649.744	0.0899	649.654	Beta	α= 0.147, β =0.997
W24	101.767	117.278	608.486	0.0973	608.389	Lognormal (3P)	α= 3.712 β =2.450 ã = 0.097
W25	42.429	70.780	314.256	0.0042	314.252	Gamma	α = 0.359 β = 118.070
W26	27.255	70.416	219.652	0.0125	219.639	Weibull	α = 0.501 β = 13.657
W27	91.585	143.094	734.291	0.0095	734.282	Beta	α= 0.180 β = 1.0715
W28	47.134	105.110	447.016	0.0474	446.016	Gamma	α = 0.201 β = 234.400
W29	51.309	96.578	476.720	0.0973	476.623	Weibull (3P)	α=0.355 β =26.945 ã = 0.051
W30	54.054	105.075	394.252	0.0077	394.244	Lognormal (3P)	α = 2.908 β = 1.508 ã = 0.007
W31	22.622	40.990	214.109	0.0167	214.092	Gamma	α= 0.304 β = 74.272
W32	26.589	46.065	203.794	0.0568	203.794	Gamma	α= 0.333 β = 79.808
W33	73.855	122.243	436.866	0.0399	436.826	Weibull	α = 0.447 β = 32.125
W34	79.361	115.372	561.097	0.0182	561.079	Gamma	α= 0.260 β = 304.190
W35	148.57	108.621	666.672	0.0144	666.677	Weibull	α = 0.371 β = 36.986
W36	128.260	163.800	789.801	0.0408	789.760	Weibull	α = 0.480 β = 21.416
W37	142.207	136.143	713.265	0.0175	713.248	Weibull	α= 0.367 β = 16.940
W38	143.770	127.144	702.751	0.0516	702.700	Gamma	α = 0.220 β = 697.280
W39	110.718	76.004	533.588	0.0077	533.581	Lognormal (3P)	α= 3.983 β = 1.668 ã = 0.007
W40	60.2381	108.428	389.787	0.0332	389.754	Reciprocal	a = 0.033 b = 167.800
W41	147.534	167.415	543.434	0.0267	543.408	Gamma	α= 0.140 β = 119.900
W42	27.723	46.721	208.356	0.0400	208.316	Beta	α = 0.198 β = 1.169
W43	23.0184	76.721	374.342	0.0003	374.341	Pareto 2	α = 0.471 β = 0.361

TABLE 4

Empirical Distribution and Best Fitted Probability Distribution of Monthly Rainfall

Month	Mean	SD	Maximum	Minimum	Range	Fitted probability distribution	Parameter values
January	24.617	0.786	4.475	3.248	1.227	Gamma	α=2.332 β=0.3405
February	27.467	1.633	3.744	1.409	2.335	Gamma	α=2.2875 β=0.4141
March	28.129	1.652	4.579	2.279	2.300	Gamma	α=2.4316 β=0.3732
April	42.644	1.411	6.3545	5.687	0.667	Gamma	α=3.0063 β=0.6946
May	71.504	1.420	11.274	9.208	2.066	Gamma	α=0.2771 β=19.9627
June	67.532	1.879	9.0680	6.856	2.211	Gamma	α=0.5332 β=10.4726
July	110.760	2.564	12.209	6.369	5.839	Gamma	α=14.6047 β=0.3889
August	196.042	8.683	30.769	9.991	20.778	Gamma	α=6.3950 β=0.9889
September	173.382	2.826	20.146	15.111	5.035	Gamma	α=4.2064 β=1.3740
October	137.436	2.807	16.012	5.200	9.811	Gamma	α=5.8513 β=0.7577
November	85.470	1.288	10.892	8.102	2.789	Gamma	α=3.1714 β=1.4764
December	35.397	0.549	4.752	4.440	0.311	Gamma	α=2.8195 β=0.4393

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 Modified Difference Method.

It is noticed that for the weekly rainfall data Gamma and the Weibull distributions were found to be the best fit for most of the weeks. In addition to that Lognormal, Pareto, Beta and Reciprocal distribution were also identified as suitable for some weeks.

For the monthly Rainfall data Gamma distribution was seen as the best fit, and the parameter values of that are detected by using the long method. Empirical distribution for weekly data and also the monthly data was recorded and it has shown that 35th standard meteorological week and the month of August has got the highest rainfall.

REFERENCES

GRAHAM, L. P., ANDRÉASSON, J. AND CARLSSON, B., 2007, Assessing climate change impacts on hydrology from an ensemble of regional climate models, model scales and linking methods - A case study on the Lule River basin. *Climate Change*, **81** : 293 - 307.

GREENWOOD, J. A. AND DURAND, D., 1960, Aids for fitting the gamma distribution by maximum likelihood. *Technometrics*, **2** : 55 - 65.

INES, A. V. AND HANSEN, J. W., 2006, Bias correction of daily GCM rainfall for crop simulation studies. *Agril. Forest Meteorol.*, **138** (1) : 44 - 53.

JOHNSON, N. L. AND KOTZ, S., 1970, Continuous Univariate Distributions, I. Houghton Mifflin Company, Boston, MA.

LARS S. HANSON AND VOGEL RICHARD, 2008, The probability distribution of daily rainfall in the United States. *Proc. of World Environmental and Water Resources Congress*, 1 - 10.

MANIKANDAN, M., THIYAGAGARAJAN, G. AND KUMAR, V. G., 2011, Probability analysis for estimating annual one day maximum rainfall in Tamil Nadu Agricultural University. *Madras Agril. J.,* **98** (1-3) : 69 - 73.

PIANI, C., HAERTER, J. O. AND COPPOLA, E., 2010, Statistical bias correction for daily precipitation in regional climate models over Europe. *Theoretical Appl. Climate*, **99** (1-2) : 187 - 192.

TAOFIK, O. D., ASIRIBO, O. E. AND AKINTOYE, N. A., 2013, On the estimation of scaled Weibull Distibution of rainfall data of South west Nigeria. *Int. J. Modern Math. Sci.*, **7** (2) : 176 - 184.

VERNIMMEN, R. R. E., HOOIJER, A., ALDRIAN, E. AND VAN - DIJK, A. I. J. M., 2012, Evaluation and bias correction of satellite rainfall data for drought monitoring in Indonesia. *Hydrol. Earth Syst. Sci.,* **16** (1) : 133 - 146.

WEILAND, S. F. C., BEEK, V. L. P. H., KWADIKJ, J. C. J. AND BIERKENS, M. F. P., 2010, The ability of a GCM-forced hydrological model to reproduce global discharge variability. *Hydrol. Earth Sys. Sci*., **14** (8) : 1595 - 1621.

(*Received :* January, 2018 *Accepted :* November, 2018)