International Journal on Emerging Technologies **11**(5): 545-557(2020)

ISSN No. (Print): 0975-8364
ISSN No. (Online): 2249-3255

Statistical Analysis and Bias Correction of GSMaP Satellite Rainfall Data for Flash Floods Modelling over the Basin of Ravi River, Pakistan

Abdullah[1], Habib-Ur-Rehman[2] and M. Ali Mirza[3]
[1]Lab Engineer, Department of Civil Engineering,
University of Management and Technology (UMT), Lahore, Punjab, Pakistan
[2]Dean/Professor, Department of Civil Engineering,
University of Engineering and Technology (UET), Lahore, Punjab, Pakistan.
[3]Lecturer, Department of Civil Engineering,
University of Engineering and Technology (UET), Lahore, Punjab, Pakistan.

(Corresponding author: Abdullah)
(Received 10 September 2020, Revised 16 October 2020, Accepted 13 November 2020)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: The basin of Ravi has a very weak network of rain gauges. This issue excels the importance of SPEs to be used extensively instead of gauge products for different hydrological investigations like flash flood modeling etc. The primary purpose of this research was to assess the SPEs and make them useful for flash flood studies by applying bias correction procedure on it. The SPEs showed a linear correlation with rain gauge products however under and overestimations in comparison with the gauge products were also observed. A comprehensive statistical analysis was carried out and multiplicative bias correction procedure was applied on SPEs individually for point precipitation of each rain gauge, then whole examined area and at the end for average basin precipitation, after categorizing them on the basis of intensity. A percentage bias of 16.75 % was observed for point precipitation and 6.30 % for average basin precipitation of the study area, which were then significantly reduced to 0.78% and 0.31% respectively after bias corrections. Flash floods modelling was performed using gauge products, biased and corrected GSMaP products for the years 2014-2016. Upstream boundary condition was provided by a flow hydrograph observed at Jassar stream gauging station. The modelling results on flash flood showed improvement because of bias correction procedure applied on GSMaP product. The results of bias correction of rainfalls using multiplicative bias correction factor showed better match with ground-based estimates. This was proved with improved values of statistical indices.

Keywords: Satellite Precipitation Estimates (SPEs), GSMaP, Bias Correction, Flash Flood Modelling, Ravi River, Pakistan.

Abbreviations: SPEs, Satellite Precipitation Estimates; GSMaP, Global Satellite Mapping of Precipitation; DEM, Digital Elevation Model; HEC-HMS, Hydrologic Engineering Centre-Hydrological Modelling System; PMD, Pakistan Meteorological Department; PST, Pakistan Standard Time; RMC, Regional Meteorological Centre; CN, Curve Number.

I. INTRODUCTION

During the years 2010-2014, Pakistan faced floods and flash floods which resulted the loss of over 1-million-acre crop lands. In 2014 heavy monsoon rains and floods in the basins of Indian eastern rivers of Jhelum, Ravi, Sutlej and Chenab induced flash floods in Punjab and other provinces of Pakistan [24]. The Basin of Ravi River is located in Punjab Province of Pakistan, containing the cities of Zone B and D where the temperature is normally ≤ 25℃ for the day time in cold season (October to March), while it is ≥ 40℃ during hot season (April to September). Generally light rainfall is observed during cold season, but extreme rain events are perceived during hot season [26]. In order to estimate flash floods, rain gauge networks existing on ground are too scarce, therefore large assumptions are demanded to plan isohytal charts over the study area [19].

For the management of freshwater resources and the prediction of heavy influence weather events like typhoons, hurricanes and heavy rains which produce landslide and heavy floods, the availability of unfailing global rainfall data and truthful time-based precipitation evaluations are necessary [11]. However, the precipitation estimation is one of the utmost problematic observational chores of climatology as the precipitation takes place occasionally and with distinct topographical and temporal unpredictability [2]. Traditional rain gauge structures yield relatively accurate point-based precipitation estimates [6, 7]. However, irregular distribution of these gauges and restricted sampling area under their coverage, produce significant complications for usefulness of their spatial extent [37]. Additionally, the deserted and far-off areas are not shielded under rain gauges [6, 7]. Furthermore, radars used for the estimation of precipitation have quantitative coverage limited to 150 Km, therefore, gauge, radar and satellite products must be combined to enhance the quality of temporal and spatial rainfall estimations [4, 6, 7].

The only standing base of data for operational hydrologic and flash flood prediction because of spatial limitations of gauge evaluations are the Satellite-built Precipitation Estimates (SPEs). Unlike ground-built rain gauges, SPEs are unprotected from huge systematic flaws and extra uncertainty bases [31]. Presently numerous satellite-built precipitation datasets are reachable which offer temporal-based and spatial-based distribution of rainfall over the area under investigation [20]. Satellite-built rainfall products estimate precipitation by means of indirect methods and thus their quantities are dissimilar in comparison to rain gauge measurements. Although remarkable progress has been accomplished in recent years for the improvement and accessibility of real-time SPEs, but still they exhibit significant inaccuracies that are essential to be corrected for any hydrological practice for example real-time or periodic flow forecasting. These uncertainties are lying because of flawed evaluation of spatial-based and temporal-based climatic factors, or the improper understanding of precipitation producing mechanisms [34].

Although, SPEs are accessible without restrictions for most of the countries on the globe, there are some limitations in these estimates, mainly due to indirect precipitation retrieving approaches. SPEs algorithms approximate precipitation by means of remote sensing techniques in which electromagnetic spectrum is used having sensors which are sensitive to visible, infrared (IR) and microwave [1]. Visible and infrared (IR) sensors are reachable on geostationary circumnavigating satellites and the precipitation data is offered by them on finer temporal scales. The precipitation products prearranged by these sensors are often unpolished as the temperature of cloud-top is meanderingly and weakly connected to rainfall [25]. Current satellites-based estimates fuse the data from different bases for example rain gauges, IR and PMW to operate the advantage of their strength. Few of free and open access SPEs are TRMM [14], TMPA [12], CMORPH [16, 38], PERSIANN and PERSIANN-CDR [30], GSMaP [18] and IMERG [13].

Only few studies have been perceived in which SPEs are analyzed, corrected and further applied for floods investigations in the Ravi basin [19]. [3] adjusted the TRMM product for Indus river basin, Pakistan by means of regression and geographical differential analysis. [21] conducted the uncertainty analysis of SPEs via couple method. [19] conducted the introductory valuation of IMERG research and IMERG real time product for the five different climate zones of Pakistan. [19] took GSMaP product, corrected it by applying bias correction procedure and went for flash flood simulation using Hydro-BEAM model.

Many studies are done using hydrological models, but now modern studies are utilizing hydraulic models for this purpose. [28] evaluated and bias corrected GSMaP_MVK product via power function and generalized additive model, in his study of flood examining in Kyushu, Japan. However, the functioning of SPEs especially GSMaP has not been conducted over Ravi river basin for different hydrological studies before.

In general, many studies can be seen worldwide in which the SPEs are analyzed for various hydrological investigations like rainfall runoff modelling, flash flood modeling, hydrological modelling and flood monitoring etc. [8] investigated GSMaP and TRMM satellite products and calibrated the aforementioned products by comparing with gauge products. The rainfall runoff simulation was executed by using LANDPINE model. [35] examined three SPEs namely TMPA 3B42, PERSIANN and CMORPH by concentrating on the capability of these SPEs to perceive low rainfall patterns which will possibly bring about the droughts. Each of these SPEs underestimated the precipitation during dry times of the year. [31] investigated the SPEs, gauge and radar products by emerging a tactic which flawlessly intermingle all the three products. For intermingling these products, the bias of each product comparative to each other was eradicated. The multiplicative bias correction factor was applied in this investigation on hourly SPEs and the results were compared with previously applied bias correction techniques. [9] examined the CMPORPH satellite product for rainfall runoff modelling, after removing the bias from SPEs HBV model was applied to conduct rainfall runoff modelling. In another research conducted by [34], two bias correction techniques namely; Quantile Mapping and the Principal Component-based technique were adopted. Later the HYMOD_DS hydrological model was used for hydrological forecasting. [22] evaluated three SPEs which were TMPA, CMORPH and CHIRPS. A comprehensive comparison was done on various temporal scales for the period of 2008-12. The topographic operated model (TOPMODEL) was implemented to see the influence of bias in SPEs on the closure of water balance.

In this research, the daily precipitation estimates of GSMaP satellite product were statistically evaluated and corrected by applying multiplicative bias correction procedure [25, 31, 32] for seven rain gauges stations. The selected region is Ravi river basin which is a very fertile plain land in Punjab, Pakistan. This kind of research is novel for this basin. The SPEs were comprehensively analyzed step by step, by considering each station's point rainfall, entire study area's point rainfall and then the average basin rainfall for the years 2014-16. After the bias correction of satellite product, the hydrological model HEC-HMS was used for modeling the flash floods generated by gauge precipitation, biased satellite precipitation and corrected satellite precipitation.

II. MATERIALS AND METHODS

A. Study Area

The area under investigation was Ravi river basin, located in almost middle part of province Punjab in Pakistan as shown in Figure 1a. It extends between 30°30'N - 32°48'N latitude ranges and 71°48'E - 75°48'E longitude ranges. The climate classification of the study area is divided into Zone B and D which includes both dry winter and rainy summer season [26].

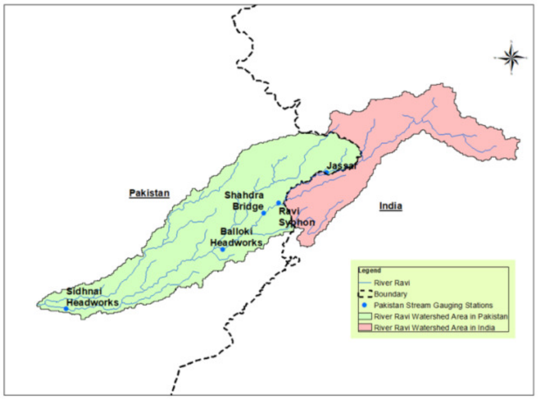

Fig. 1 (a) The total catchment area of Ravi Basin delineated by means of SRTM 90 m DEM.

When the basin was outlined by means of 30 m Digital Elevation Model (DEM) in ArcGIS, total area of Ravi basin was perceived as 52,800 Km^2 (Figure 1a), out of which 33264 Km^2 lies in Pakistan and the rest of the area is included in Indian territory. In this study, only Pakistani basin area of river Ravi was considered.

By exercising Thiessen Polygons algorithm in ArcGIS, it became evident that the seven rain gauge stations (Fig. 1b), which were covering the study area, shielded 31488.44 Km^2 area of Pakistani Ravi basin. The remaining area of Pakistani Ravi basin is supposed to be under the coverage of Indian rain gauges. However, while working on average basin precipitation only the area estimated by Thiessen Polygons algorithm was compared for gauges and GSMaP precipitation.

The cities inside and around the basin area are Faisalabad, Lahore, Multan, Toba Tek Singh, Sahiwal, Gujranwala and Okara. The Ravi river flows through the streams of these towns and finally discharges into Chenab river near Multan [19]. The summer season (April to September) is relatively rainier as compared to winter season (October to March) which is drier. During summer heavy rainfall events occur in the months of July, August and September [26]. The soil is quite fertile, good for farming and classified as alluvial soil [27].

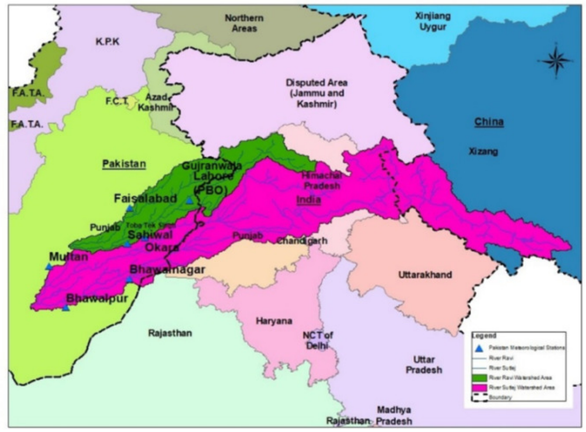

Fig. 1 (b) Study area (Green Color) showing the location of Rain Gauging Stations.

B. Gauge built and Satellite built rainfall datasets

The daily rainfall data with spatial resolution 0.1° × 0.1° for three years 2014 to 2016 was downloaded from Japan based GSMaP's official website named as Jaxa Global Rainfall Watch. The locations of rain gauge stations were simply entered in GSMaP's website one by one, and the results were collected as a point value of precipitation in the form of pixel value. The observation time in Pakistan for rain gauges is set as 08:00 AM by Pakistan Meteorological Department (PMD). GSMaP follows UTC which is 5 hours behind Pakistan Standard Time (PST). So, the observation time was set as 03:00 to 03:00 for GSMaP to match the observation time. The link to the GSMaP official website is given as;

https://sharaku.eorc.jaxa.jp/GSMaP/

Table 1 shows the specifications of GSMaP dataset.

The daily rainfall data for seven rain gauge stations was obtained from Regional Meteorological Centre (RMC), PMD, Lahore, Pakistan for years 2014 to 2016. The particulars related to these rain gauge stations are hereby listed in Table 2. Fig. 1b is showing the location of rain gauging observatories covering the study area. The locations of rain gauge stations listed in Table 2 were collected from the website of PMD.

Table 1: Details of SPE product [25, 28].

Characteristics of GSMaP product	
Sensors	Microwave and Infrared Sensors
Spatial Resolution	0.1° × 0.1° and 0.25° × 0.25° grids
Temporal Resolution	Hourly and Daily
Start Date	March 2000
Coverage	60°N to 60°S

Table 2: Rain Gauge Stations Details.

S.No.	Name of Station	Longitude (X)	Latitude (Y)
1.	Faisalabad	73°7'60"	31°25'60"
2.	Lahore	74°19'27.89"	31°32'34.13"
3.	Multan	71°25'30.89"	30°11'50.04"
4.	Toba Tek Singh	72°46'60"	30°58'60"
5.	Sahiwal	73°10'00"	30°38'60"
6.	Gujranwala	74°20'60"	32°10'26.90"
7.	Okara	73°25'60"	30°48'00"

C. Description of Hydrological Model and its Execution

Flash flood simulations were modelled using HEC-HMS for daily rainfall data of years 2014 to 2016. The model was set to function by fixing the loss method as SCS-Curve Number method, runoff estimation method as SCS-Unit Hydrograph method and flow routing method as Muskingum method, were adopted. By bearing in mind the study area, the lag time was taken as 24 hours, which is the time taken by rain water to reach the river channel. The Curve Number (CN) was taken as 78 as the Punjab soil is classified as agricultural alluvial soil, booked in Hydrological Soil Group B [23, 27]. The time interval for control specification was taken as 1 day, owing to the availability of daily rainfall data. The Upstream boundary condition was provided by a flow hydrograph observed at Jassar stream gauging station. The flood simulation results were attained in the form of graphs, displaying peak discharge for daily rainfall values (Fig. 9). The governing equations used in HEC-HMS are;

$$I_a = S\ (\text{in mm})\ \times\ 0.2 \quad (1)$$

$$S = \frac{1000}{CN} - 10 \quad (2)$$

$$T_{lag} = \frac{L^{0.8} \times (S+1)^{0.7}}{(1900 \times Y^{0.5})} \quad (3)$$

$$Y = \frac{(\text{Higher Elevation} - \text{Lower Elevation})}{L} \quad (4)$$

Where I_a is initial abstraction which is the loss of water due to evaporation, infiltration and interception. This is calculated in mm and is dependent on Curve Number (CN). *S* is maximum retention in watershed. *CN* is Curve Number. T_{lag} is the lag time i.e. the time taken by water to reach the river channel. *Y* is the slope and *L* is the length between higher and lower elevation. The flow diagram showing the brief methodology in steps is attached (Fig. 2).

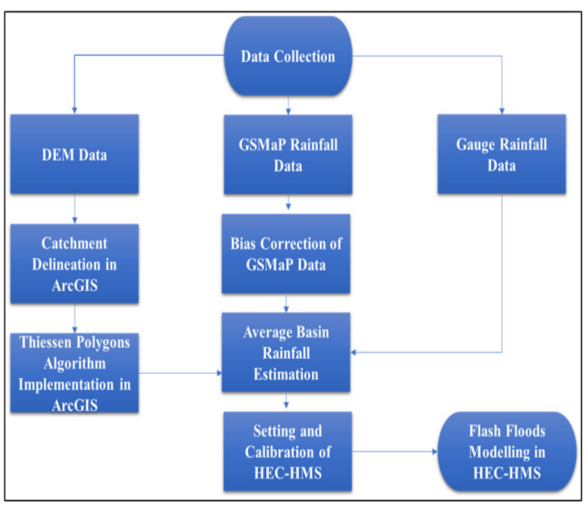

Fig. 2. Flow Diagram Showing Research Methodology.

D. Comparison of Gauge and GSMaP precipitation datasets

The gauge and GSMaP rainfall products were compared primarily for the assessment of GSMaP product, over the Ravi river basin. To make the spatial resolution of gauge rainfall dataset harmonious with the GSMaP dataset, it was interpolated to 0.1° × 0.1° spatial resolution grid by applying the algorithm for generation of programmed Thiessen Polygons [10]. This was done by assigning the equal rainfall value to the pixels neighboring the gauge station. This interpolation practice essentially depends upon the concentration of gauge networks in the study area, and it can additionally introduce further errors in rainfall dataset. The geostatistical models which are more comprehensive, require denser network of rain gauge stations for characterization of error formation. One good example for such models is Kriging [25].

The comparison between the two datasets was made at different scenarios i.e. the comparison between the point precipitation of rain gauge and pixel value of GSMaP for each station, the same comparison by considering the point/pixel values for complete study zone and finally the comparison between average basin precipitations estimated using Thiessen Polygon algorithm, for gauge and GSMaP datasets. All these comparisons were established on the basis of daily rainfall records. The heavy rainfall events were given special attention during the comprehensive statistical analysis, as these events are crucial in flash flood modelling or simulations.

E. Statistical Parameters

The comparison between the gauge and GSMaP rainfall datasets was established by evaluating four statistical parameters [15, 19, 25] which are; Bias (B), Relative Bias (RB), Root Mean Square Error (RMSE) and Correlation Coefficient (CC), shown in Table 3. B represents the average value of difference between rain gauge rainfall values and SPEs. B can be positive or negative, if B is positive it represents overestimation and if it is negative it indicates underestimation. RB shows the systematic error present in SPEs. RB also shows positive and negative values, depicting overestimation and underestimation accordingly. RMSE is the most commonly applied method and gives absolute value of average error and is sensitive to larger values of error [5, 29]. CC is the agreement between the rain gauge values and SPEs.

Further, there are some statistical parameters which were assessed for consistency and categorial authentication of datasets i.e. Probability of Detection (POD), False Alarm Ratio (FAR) and Critical Success Index (CSI). POD is the rate of hit values i.e. the fraction of accurately sensed rainfall events and ranges from zero to 1. FAR states the rainfall values which were incorrectly detected. Lastly, CSI is the segment of precipitation events appropriately perceived by GSMaP. Statistical parameters with their equations and best values are cataloged in the Table 3.

Table 3: Statistical Parameters used in the Comparison and Evaluation.

Statistical Parameters	Equations	Best Values
Bias/Mean Error (B/ME)	$B = \frac{1}{n}\sum_{i=1}^{n}(S_i - G_i)$	0
Relative/Percent Bias (RB/PB)	$RB = \frac{\sum_{i=1}^{n}(S_i - G_i)}{\sum_{i=1}^{n} G_i} X\ 100\%$	0
Root Mean Square Error (RMSE)	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(S_i - G_i)^2}$	0
Correlation Coefficient (CC)	$CC = \frac{\sum_{i=1}^{n}(G_i - \bar{G})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^{n}(G_i - \bar{G})^2}\sqrt{\sum_{i=1}^{n}(S_i - \bar{S})^2}}$	1
Probability of Detection (POD)	$POD = \frac{A}{A + C}$	1
False Alarm Ratio (FAR)	$FAR = \frac{B}{A + B}$	0
Critical Success Index (CSI)	$CSI = \frac{A}{A + B + C}$	1

Where, n is the quantity of rain gauge or GSMaP values; S_i is the satellite-built estimations and G_i is rain gauge-built amounts. A is "hit" i.e. accurately measured rainfall events by GSMaP in comparison with rain gauge. B is "false alarm" i.e. when no rainfall is happened actually, but GSMaP estimates the rainfall. C is "miss" i.e. when rainfall occurs but GSMaP show zero value of rainfall [28].

F. Bias Correction of GSMaP Precipitation Product

The precise simulations for flash flood are possible if we are using corrected satellites estimates and matching them with simulation results of rain gauge precipitation values [36]. Therefore, it was examined in hydrological model HEC-HMS that whether the corrected GSMaP rainfall product, along with gauge estimates, is giving improved flood simulations or not. The GSMaP data was corrected by means of simple multiplicative bias correction factor because of the presence of scarce system of rain gauge stations [31]. In a study conducted by [32] additive and multiplicative error removal techniques were devised, and proposed the practice of multiplicative bias correction factor for SPEs present in the shape of daily rainfall information. The bias correction was performed by employing the monthly-based bias rectification factor for the correction of hourly rainfall data, which was then converted to daily precipitation data [25]. The equation for multiplicative bias factor is given below;

$$Bias\ Correction\ Factor\ (BF) = \frac{Gauge\ Precipitation\ (G_i)}{GSMaP\ Precipitation\ (S_i)} \quad (5)$$

$$GSMaP_{Corrected}(S_{i_{Corr}}) = Bias\ Correction\ Factor\ (BF) \times GSMaP\ Precippitation(S_i) \quad (6)$$

The bias correction was done very carefully by categorizing the incorrect GSMaP rainfall data into different ranges, and the values where the error was more than 50% a factor of 2 was utilized for its bias correction [25]. In this way, most of the data points reached nearer to 45° reference line in scatter plot. For the bigger rainfall values more deviation from the reference line was detected.

G. Calibration and Validation of HEC-HMS

Calibration of HEC-HMS model was executed in order to acquire the peak discharge and the time of that peak discharge both manually and by using software. The Shuffled Complex Evaluation (SCE) algorithm was used for the calibration of hydrological model [25]. However, mainly the volume of discharge was simulated in this research. Therefore, the model calibration was attained by taking into account the volume of simulated discharge. Initially software-based calibration was done by considering the peak flow months (July, August and September) for year 2014 to 2016, which was then compared with manual calibration. The model's sensitivity was improved by applying SCE algorithm again and again. The model functioning was evaluated by the parameters CC, RMSE, B and RB respectively.

The validation of model was carried out by taking into account the same indices as that used for calibration but different time was chosen i.e. April, May and June 2014 to 2016. The detailed discussion on the performance of HEC-HMS model is given in results and discussion section.

III. RESULTS AND DISCUSSION

A. Point Precipitation Comparison

The comparison for point precipitations of gauge and GSMaP products was constructed expansively, by taking all the gauge stations into account one after another and then considering the entire study area under examination. The statistical parameters were evaluated and analyzed broadly, before and after applying the bias correction technique on GSMaP dataset. The evaluated parameters were B, RB, RMSE and CC. The categorial authentication indices were determined only for biased GSMaP products in comparison with gauge estimates. The detailed results and discussions are given in below sections. The statistical parameters and categorial authentication factors are listed in Table 4, 5 and Fig. 8.

Before the operation of multiplicative bias removal factor for Faisalabad station, the two concerned datasets were indicating the B value of 0.74 which is showing overestimation i.e. GSMaP overestimated the rainfall as compared to Gauge, RB value of 14.95% which is also an overestimation signal, RMSE of 78.64 which was highest among all stations under study area, the bigger value is also representing the sensitivity of RMSE to larger errors [28]. CC came to be 0.20 which shows very weak connection between gauge and GSMaP datasets.

However, the POD was 0.73 which is acceptable but not upright, FAR was 0.70 which illustrates that there were many events for which GSMaP estimated some rainfall value but actually there wasn't any precipitation.

CSI value was spotted 0.27 i.e. out of 1 part only 0.27 was the accurately assessed data (Table 4 and 5).

$$Bias\ Factor = \frac{G_i}{S_i} \quad (7)$$

$$S_{i\_Corr} = BF \times S_i \quad (8)$$

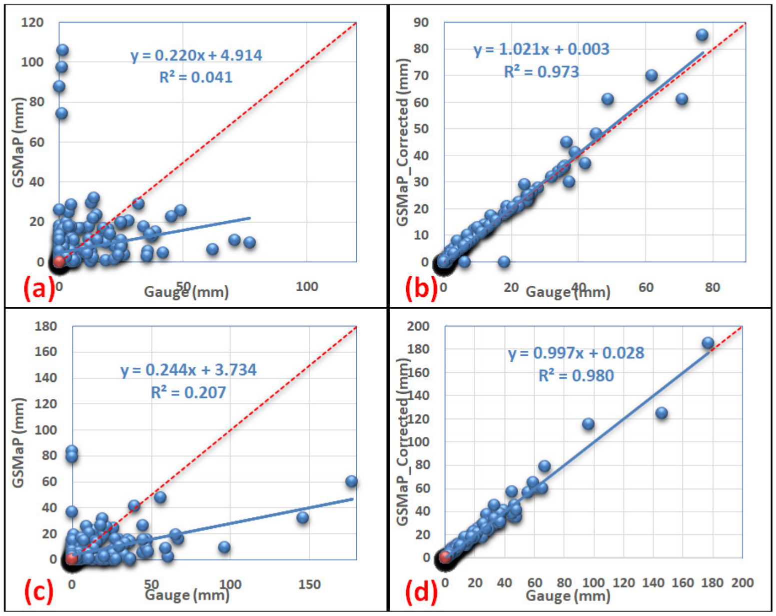

Fig. 3. Linear Scatter Plots for Daily Point Precipitations **(a)** For Faisalabad station involving Gauge and GSMaP products **(b)** For Faisalabad station involving Gauge and Corrected GSMaP product **(c)** For Lahore station involving Gauge and GSMaP products **(d)** For Lahore station involving Gauge and Corrected GSMaP product.

Later, when GSMaP was rectified through bias correction procedure (Eqns. 7 and 8), B value reduced from 0.74 to 0.11, RB shrunk from 14.95% to 2.23%, RMSE lowered from 78.64 to 1.76 and CC improved from 0.20 to 0.98 changing the relationship between products from weak to strong [34]. The bias correction mechanism showed much improvement in GSMaP product (Fig. 8). Consequently, this improved the flash flood simulation results [25]. The scatter plots before and after the implementation of bias removal method are clearly representing the behavior of data before and the correction. The improvement in the linear regression can obviously be perceived in given plots (Fig. 3a and 3b).

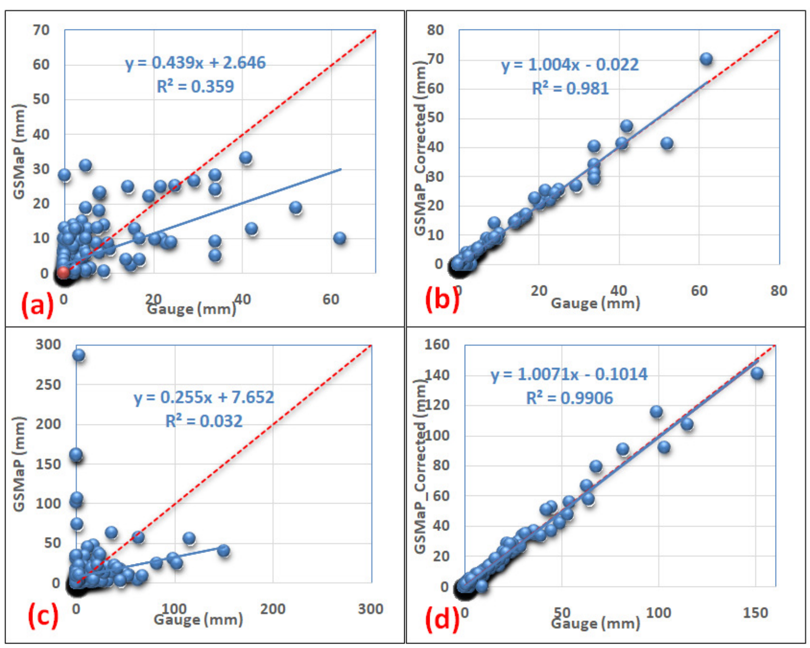

Fig. 4. Linear Scatter Plots for Daily Point Precipitations **(a)** For Multan station involving Gauge and GSMaP products **(b)** For Multan station involving Gauge and Corrected GSMaP product **(c)** For Gujranwala station involving Gauge and GSMaP products **(d)** For Gujranwala station involving Gauge and Corrected GSMaP product.

For Lahore station as evident from Fig. 3b and 3c, initially B, RB, RMSE and CC were -2.20, -30.93%, 40.88 and 0.46 respectively (Table 4). The value of B and RB are negative in this case, which confirms underestimation of rainfall magnitude by GSMaP as compared to rain gauge. RMSE is 40.88 which is quite bigger value because of large errors in the data, but lesser than that of Faisalabad station. CC is 0.46 which indicates that almost half of the gauge and GSMaP data is accurately correlated for this station. Remaining parameters POD, FAR and CSI were 0.38, 0.48 and 0.28 (Fig. 8).

The POD is low as compared to Faisalabad station, which signifies that only 38% rain events were accurately sensed in this case. FAR is quite elevated which is also not fine, as it means that almost half of the times GSMaP measured rainfall when actually there was not any rainfall there. And lastly, CSI is almost same like that of Faisalabad station. These statistics were improved after the bias correction was carried on i.e. B, RB, RMSE and CC changed to 0.01, 0.19%, 1.11 and 0.99 respectively. The correlation between the gauge and corrected GSMaP got pretty solid, making the flood modelling more accurate and reliable. The bias correction of GSMaP rendered it to be suitable for hydrological modelling, which is very evident from the scatter plots in Fig. 3c and d.

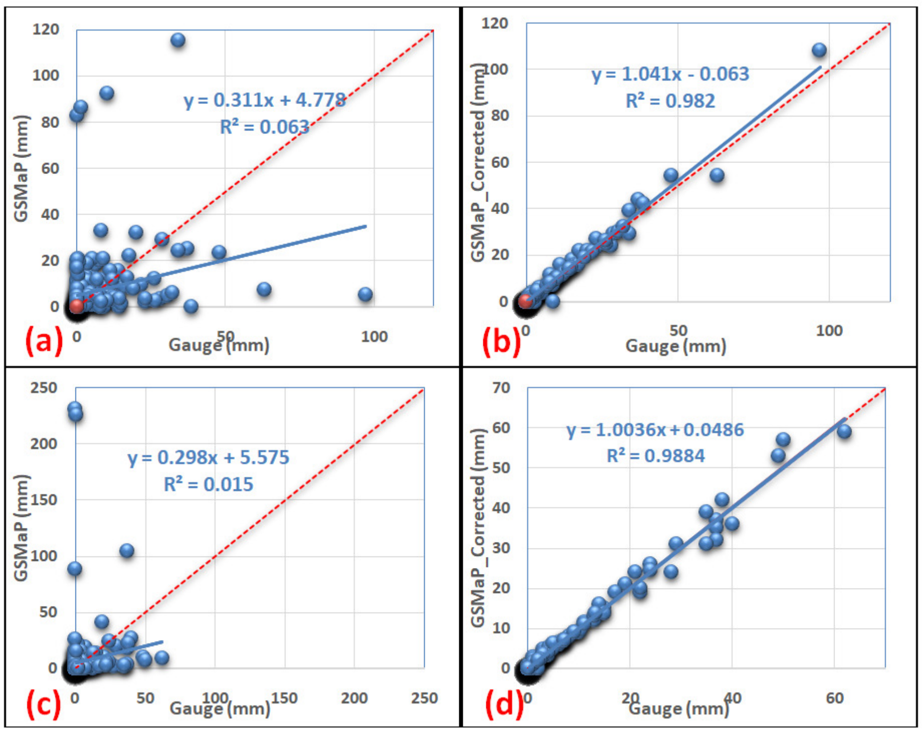

Fig. 5. Linear Scatter Plots for Daily Point Precipitations **(a)** For Toba Tek Singh station involving Gauge and GSMaP products **(b)** For Toba Tek Singh station involving Gauge and Corrected GSMaP product **(c)** For Sahiwal station involving Gauge and GSMaP products **(d)** For Sahiwal station involving Gauge and Corrected GSMaP product.

In case of Multan station, the indices of B, RB, RMSE and CC developed from 0.68 to -0.01, 21.29% to -0.24%, 61.48 to 0.01 and 0.60 to 0.99 respectively (Table 4 and 5, Fig. 8). Here if we notice the development in the GSMaP data, the B and RB have become almost negligible. RMSE was second highest in this case as compared to other stations, which is also reduced to almost zero. Fig. 4a and 4b are showing the behavior of data related to Multan station. This station's data was having a highest value of CC with GSMaP as compared to other station, which is also further improved by bias correction. Moreover, the indices associated to categorial authentication i.e. POD, FAR and CSI were sensed 0.64, 0.78 and 0.19 respectively (Table 4). The improvement in the GSMaP product for Multan station can be analyzed from above plots. The bias removal ultimately improved the results of flash flood modelling in HEC-HMS.

For Gujranwala gauge station, when the statistical analysis was done before the bias correction the statistical indices i.e. B, RB, RMSE and CC came to be 2.15, 31.49%, 40.58 and 0.18 respectively (Figg. 8). The values of B and RB are positive in this scenario, indicating the overestimation by GSMaP. RMSE is also very high as the large errors are also present in this case. The two datasets are very weakly correlated. If the GSMaP is used for flood modeling with these uncertainties present in it, it will generate very faulty results. The indices POD, FAR and CSI were 0.70, 0.74 and 0.23 correspondingly (Table 4). When the multiplicative bias correction factor was applied, the better values of statistical indices were obtained i.e. B as -0.05, RB as -0.73%, RMSE as 0.94 and CC as 0.99 (Figure 8). This is how the corrected GSMaP product can give us far better results of flash flood in the study area.

The parameters B, RB, RMSE and CC, in case of TT Singh station, were 1.39, 31.10%, 22.20 and 0.25 respectively (Fig. 8). B and RB being positively high representing overvalued rainfall by GSMaP. RMSE is though relatively low but still not satisfactory. CC is 0.25 which also not reasonable. These parameters are undoubtedly endorsing the improvement in GSMaP dataset for this station too. Categorial validation indicators i.e. POD, FAR and CSI stood 0.69, 0.73 and 0.24 respectively, demonstrating the necessity of bias correction of GSMaP dataset (Fig. 5a and 5b). Improved B, RB, RMSE and CC were achieved as 0.12, 2.76, 1.97 and 0.99 respectively, rendering the GSMaP product reliable for flash flood modelling (Table 4 and Table 5).

Table 4: Statistical categorial parameters before the implementation of bias correction for various scenarios.

S.No.	Statistical Parameters	Faisalabad	Multan	Lahore	Gujranwala	TT Singh	Sahiwal	Okara	Whole Study Area	Average Basin Precipitation
1.	*CC*	0.20	0.60	0.46	0.18	0.25	0.13	0.11	0.20	0.31
2.	*B*	0.74	0.68	-2.20	2.15	1.39	2.52	1.79	0.90	0.21
3.	*RB*	14.95	21.29	-30.93	31.49	31.1	61.49	31.61	16.75	6.30
4.	*RMSE*	78.64	61.48	40.88	40.58	22.2	38.78	27.77	39.95	4.95
5.	*POD*	0.73	0.64	0.38	0.70	0.69	0.82	0.76	0.62	0.77
7.	*FAR*	0.70	0.78	0.48	0.74	0.73	0.76	0.77	0.72	0.57
8.	*CSI*	0.27	0.19	0.28	0.23	0.24	0.23	0.21	0.24	0.38

Table 5: Statistical categorial parameters after the implementation of bias correction for various scenarios.

S.No.	Statistical Parameters	Faisalabad	Multan	Lahore	Gujranwala	TT Singh	Sahiwal	Okara	Whole Study Area	Average Basin Precipitation
1	CC	0.98	0.99	0.99	0.99	0.99	0.99	0.98	0.99	0.99
2	B	0.11	-0.01	0.01	-0.05	0.12	0.06	0.08	0.04	0.01
3	RB	2.23	-0.24	0.19	-0.73	2.76	1.55	1.49	0.78	0.31
4	RMSE	1.76	0.01	1.11	0.94	1.97	0.98	1.31	1.87	0.25

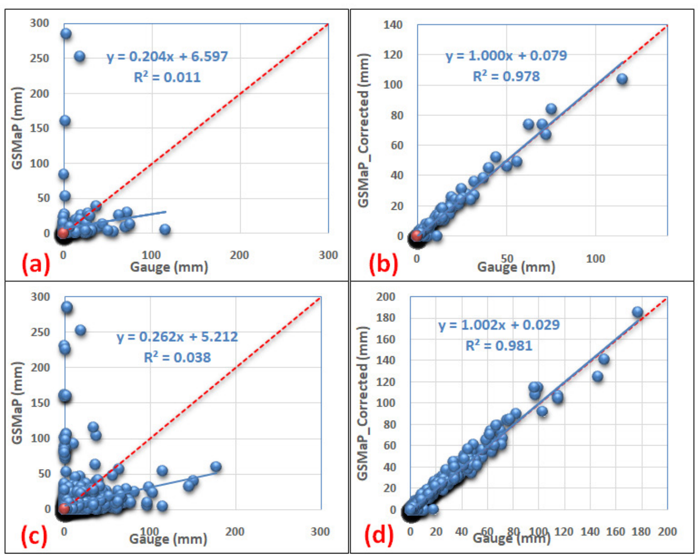

Fig. 6. Linear Scatter Plots for Daily Point Precipitations **(a)** For Okara station involving Gauge and GSMaP products **(b)** For Okara station involving Gauge and Corrected GSMaP product **(c)** For entire study area involving Gauge and GSMaP products **(d)** For entire study area involving Gauge and Corrected GSMaP product.

Statistical analysis prior the implementation of bias correction operation delivered the values of indices B, RB, RMSE and CC as 2.52, 61.49%, 38.78 and 0.13 respectively for Sahiwal station (Fig. 8). Both B and RB are positive indicating the overrating of rainfall by GSMaP. Moreover, RB in this case highest as compared to other stations covering the study area. RMSE is relatively low as compared to other observatories, but needs rectification. POD, FAR and CSI were 0.82, 0.76 and 0.23 correspondingly (Table 4 and 5). Multiplicative bias correction procedure dispensed the upgraded indices B, RB, RMSE and CC as 0.06, 1.55%, 0.98 and 0.99 (Fig. 5c and 5d), confirming the reliable productivity of flash flood modelling in the hydrologic model (Fig. 9).

Similar to other gauge stations, covering the area of interest, the GSMaP precipitation product for Okara also station contains plenty of uncertainties. The analysis of statistical indicators yielded the values of B, RB, RMSE and CC as 1.79, 31.61%, 27.77 and 0.11 respectively

(Fig. 8). The statistical indicators are plainly illustrating the crudeness of GSMaP dataset in this case. The indices POD, FAR and CSI were 0.76, 0.77 and 0.21 respectively (Table 4). It has been observed that FAR is usually high for all stations. CSI is also low in this case. The satellite-built dataset needs ultimate refinement, if it is to be used further in hydrological investigations. When bias eradication was implemented, the significant amendment in the statistical pointers was realized i.e. B, RB, RMSE and CC were developed to 0.08, 1.49%, 1.31 and 0.99 correspondingly (Fig. 6a and 6b). In this way the station wise point precipitation for each gauge station was bias corrected, so that the analysis at temporal and spatial scales can be established excellently.

For the extraction of more dependable results in the form of statistical indicators, the point precipitations of both gauge and GSMaP for complete study area was utilized. The statistical analysis for point precipitation expanded in that way. The parameters B, RB, RMSE and CC were obtained as 0.90, 16.75%, 39.95 and 0.20 respectively, similar to results which [19] achieved for IMERG. The parameters are undoubtedly demonstrating the rawness of GSMaP dataset. While on the other hand the parameters POD, FAR and CSI were 0.62, 0.72 and 0.24 correspondingly (Fig. 8). These results also showed agreement with study conducted by [28]. Though the results are demanding the rectification of GSMaP product, but we can say that these statistical indicators are more dependable as compared those for each station. When the same multiplicative bias correction factor was implemented in this scenario, the satellite-built rainfall products showed much improvement. The parameters B, RB, RMSE and CC improved to 0.04, 0.78%, 1.87 and 0.99 respectively (Figure 6c and 6d). The improvement in SPEs is evidently allowing the use of GSMaP dataset for hydrological modelling along with gauge datasets. Figure 6e is showing the mean annual rainfall of satellite product at various gauge stations and also for average basin precipitation.

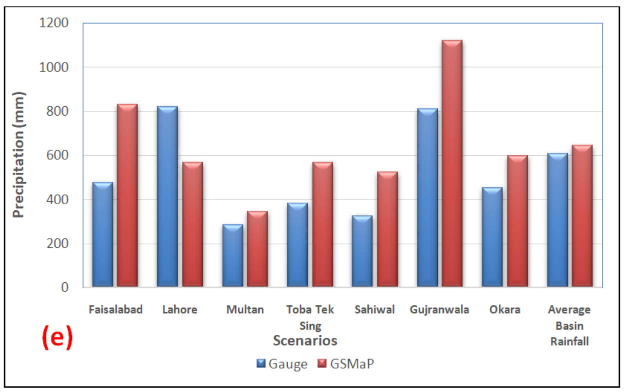

Fig. 6 (e) Mean Annual Gauge and GSMaP Rainfalls for all Stations of the Study Area.

In the coming section, the comparison between gauge and GSMaP datasets on spatial scale is also comprehensively discussed in detailed.

B. Comparison for Average Daily Basin Precipitation

In this section, the average daily basin precipitation, estimated by means of Thiessen Polygons Algorithm in ArcGIS, was compared and bias corrected. The bias correction for average basin rainfall is quite essential, as the flash flood modelling was actually performed using average basin precipitation. The values of statistical factors B, RB, RMSE and CC were 0.21, 6.30%, 4.95 and 0.31 respectively (Fig. 8), confirmed by [19] at regional scale for IMERG. These B, RB and RMSE are lowest in case of average precipitation comparison, as compared to point precipitations. This shows that the efficiency of GSMaP is quite better in measuring average basin rainfall as compared to point rainfall. The CC value is 0.31 which shows the satellite dataset still needs bias correction. The categorial confirmation indices POD, FAR and CSI were 0.77, 0.57 and 0.38 respectively. These indices are also batter in this scenario as compared to point precipitation case of each station. However, the bias correction procedure rendered, the GSMaP based average basin rainfall, advantageous for flash flood estimation in HEC-HMS, along with gauge based average basin precipitation.

The adjusted GSMaP dataset yielded the indices B, RB, RMSE and CC as 0.01, 0.31, 0.25 and 0.99 respectively (Fig. 7a and 7b). After bias correction, the GSMaP based average basin rainfall was used in HEC-HMS, and the results were very close to gauge-based flash flood values.

The functioning of HEC-HMS model for gauge, GSMaP and Corrected GSMaP products are discussed in detail in coming section.

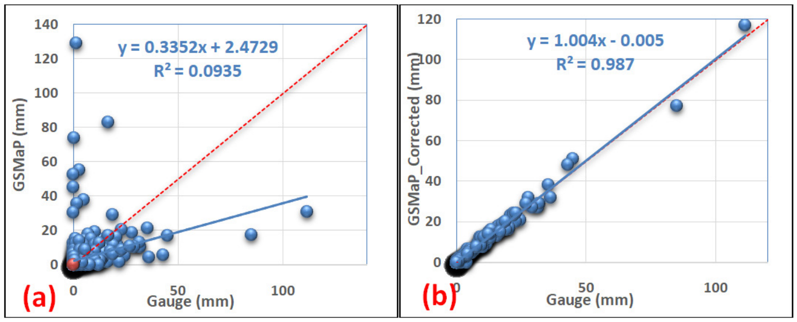

Fig. 7. Linear Scatter Plots for Average Basin Precipitations **(a)** Involving Gauge and GSMaP Products and **(b)** Involving Gauge and Corrected GSMaP Products.

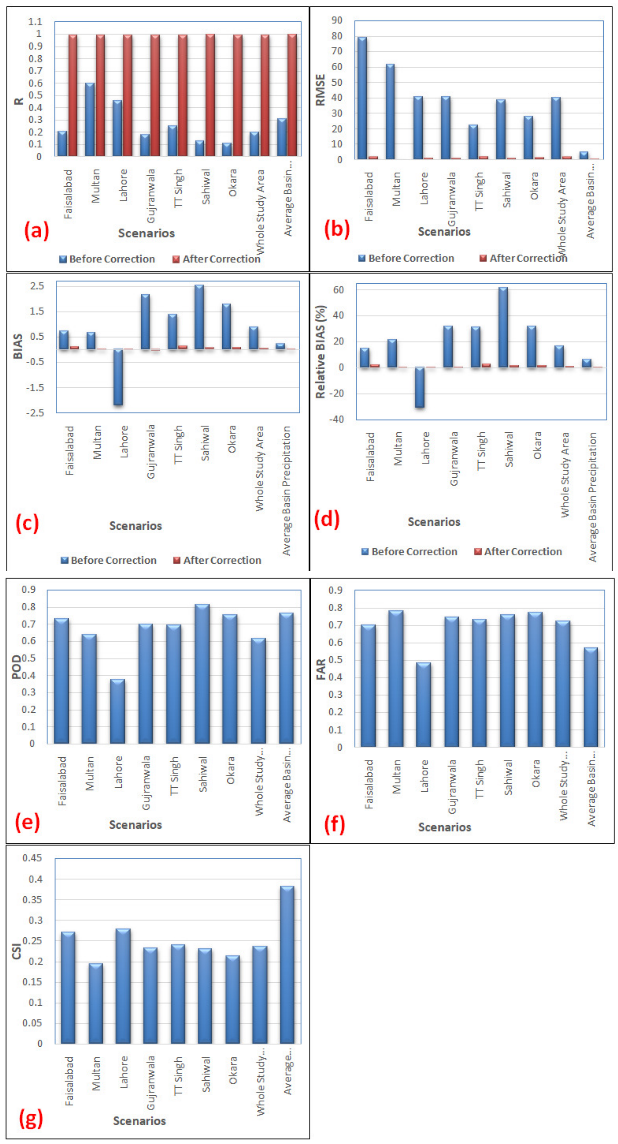

Fig. 8. Graphical representation of statistical parameters **(a)** Correlation Values for Various Scenarios Before and After the Bias Removal **(b)** RMSE Values for Various Scenarios Before and After the Bias Removal **(c)** Bias (B) Values for Various Scenarios Before and After the Bias Removal **(d)** Relative Bias (RB) Values for Various Scenarios Before and After the Bias Removal **(e)** POD Values for Various Scenarios Before the Bias Removal **(f)** FAR Values for Various Scenarios Before the Bias Removal **(g)** CSI Values for Various Scenarios Before the Bias Removal.

C. Modeling of Flash Floods over Ravi River Basin

Fig. 9 (a) is showing the flash floods graph, when the average basin rainfall by gauge dataset was used. The results from the model showed a peak discharge of 25695.6 m^3/s generated on September 06, 2014. Similarly, the Fig. 9b is representing the flash floods graph for GSMaP based average basin precipitation, in which the satellite-based precipitation was used for hydrological modeling. The results from HEC-HMS delivered the peak discharge of 24110 m^3/s on March 30, 2015. The model results for the two datasets are evidently showing a huge difference between the two products.

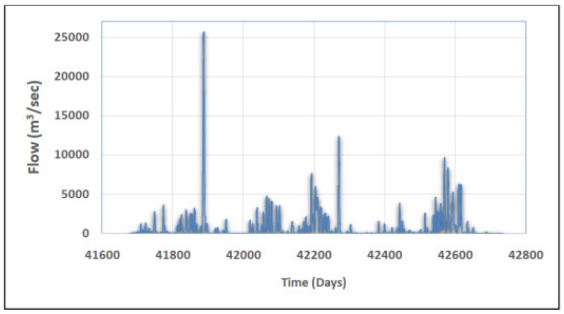

Fig. 9 (a) Flash Floods Graph for Gauge Dataset Generated by HEC-HMS.

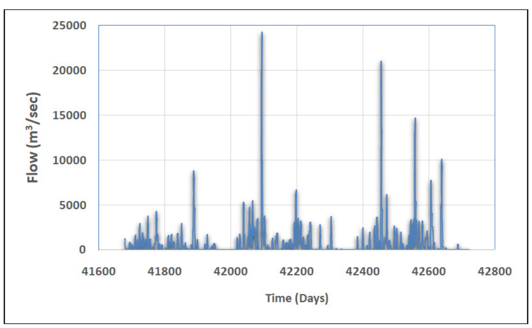

Fig. 9 (b) Flash Floods Graph for GSMaP Dataset Generated by HEC-HMS.

When the corrected average basin rainfall was used in the model the results showed a major improvement, Fig. 9c is clearly demonstrating the improvement. The Fig. 9c is depicting the flash floods modeling results in the form of a graph, in which the peak discharge was yielded as 26765.7 m^3/s on September 05, 2014. Still, there is a bit difference between the values of peak discharge and date of occurrence, however, this difference can be neglected based on the improvement exhibited by bias correction model in GSMaP product.

The GSMaP dataset was giving the peak discharge on March 30, 2015, which totally non-realistic. The gauge and corrected GSMaP are much closer in this scenario. When scatter plots were extracted (Fig. 9d and 9e), it became clear that firstly there were many differences between the flood's values of two products. Moreover, the corrected GSMaP product showed a significant improvement in the results. The statistical analysis showed that small difference in rainfall values of Gauge and GSMaP datasets delivered a bigger difference in flood's values. The B value was seen as 100.21, RB value as 9.83%, RMSE as 2320.10 and CC was 0.42 in flood values. The results are depicting major inaccuracies. The bias corrected GSMaP product improved these parameters meaningfully (Fig. 9e).

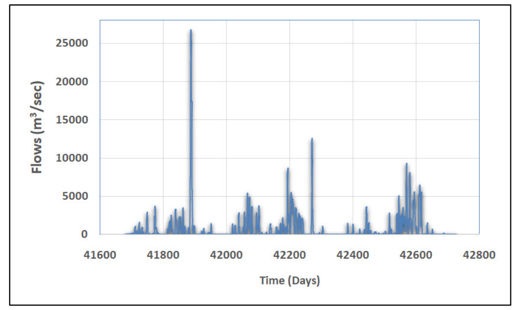

Fig. 9 (c) Flash Floods Graph for Corrected GSMaP Dataset Generated by HEC-HMS.

There are possibilities for uncertainties in model results because of bias correction technique and also because rain gauges sometimes show underestimation and overestimation as compared to satellite. There can be many reasons for error in rain gauges data like wind influences, losses due to wetting, effect of evaporation and splash [25].

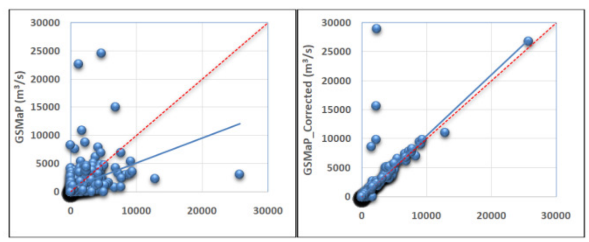

Fig. 9 (d) Flash Floods Comparison between Gauge and GSMaP **(e)** Flash Floods Comparison between Gauge and Corrected GSMaP.

IV. CONCLUSION

The basin of Ravi has a very weak network of rain gauges. This issue excels the importance of satellite-built rainfall products, to be used extensively instead of gauge products for different hydrological investigations like flash flood modeling etc. The primary purpose of this research was to assess the SPEs and make them useful for flash flood studies by applying bias correction procedure on it. A multiplicative bias correction approach was devised to adjust the GSMaP product to make them beneficial.

The investigations directed us to the conclusion that the GSMaP, in our case, mostly overestimates the rainfall in comparison with rain gauge, as a function of rainfall intensity and season of the year. It also became evident that in case of Lahore station, the GSMaP underestimated the mean annual rainfall, which may be due to presence of high rainfall intensities and variability of climate as compared to other stations. It was also found that bigger bias is observed because of the heavy intensity rainfall. This can be said as the effect of topography.

Results of bias correction of rainfalls using multiplicative bias correction factor showed better match with ground based observed rainfalls. This was proved with the improved values of statistical indices.

However, if there is an error in rainfall, the error it also yields in flash floods. Therefore, the results from HEC-HMS are clearly demonstrating that the peak discharges of Gauge and GSMaP were very much different and the month of peak was also different, which were corrected by the help of bias correction procedure. If the errors are taken into account, the GSMaP product are very much serviceable than gauge stations which are very less in number.

The results can be more refined, if hourly or three hourly rainfall estimates are used for bias correction of SPEs over Ravi Basin, as the removal of error at smaller scale will enhance the quality of larger scale SPEs. Moreover, the gauge products are not always reliable, as they may also contain error due to many reasons discussed earlier. It would be better if some third source, like radar product, is also considered for intercomparison.

In this study only SPEs from GSMaP were analyzed, bias corrected and discussed in comparison with gauge products. So, to achieve more accurate and beneficial SPEs, other satellite-built products should also be studied likewise for Ravi Basin.

V. FUTURE SCOPE

This study will enable the researchers to get familiar with the possible bias in GSMaP and the reasons related it. It will help in getting correct estimates of SPEs for their use in hydrological purposes. This will also give them idea about the presence of error in gauge-based products also.

ACKNOWLEDGEMENTS

The authors are thankful to Professor Muhammad Masood for providing the daily rainfall estimates of seven rain gauging stations covering the study area.

Conflict of Interest. There was no conflict of interest between authors or relevant organizations. The article is submitted by the mutual consent of all authors.

REFERENCES

[1]. Bhattacharya, B. K., Mallick, K., Patel, N. K., & Parihar, J. S. (2010). Regional clear sky evapotranspiration over agricultural land using remote sensing data from Indian geostationary meteorological satellite. *Journal of Hydrology*, *387*(1-2), 65-80.

[2]. Board, S. S., & National Research Council. (2007). *Earth science and applications from space: national imperatives for the next decade and beyond*. National Academies Press.

[3]. Cheema, M. J. M., & Bastiaanssen, W. G. (2012). Local calibration of remotely sensed rainfall from the TRMM satellite for different periods and spatial scales in the Indus Basin. *International Journal of Remote Sensing*, *33*(8), 2603-2627.

[4]. Chen, D., Achberger, C., Räisänen, J., & Hellström, C. (2006). Using statistical downscaling to quantify the GCM-related uncertainty in regional climate change scenarios: a case study of Swedish precipitation. *Advances in Atmospheric Sciences*, *23*(1), 54-60.

[5]. Chu, T. W., & Shirmohammadi, A. (2004). Evaluation of the SWAT model's hydrology component in the piedmont physiographic region of Maryland. *Transactions of the ASAE*, *47*(4), 1057.

[6]. Feidas, H. (2010). Validation of satellite rainfall products over Greece. *Theoretical and Applied climatology*, *99*(1-2), 193-216.

[7]. Feidas, H., Kokolatos, G., Negri, A., Manyin, M., Chrysoulakis, N., & Kamarianakis, Y. (2009). Validation of an infrared-based satellite algorithm to estimate accumulated rainfall over the Mediterranean basin. *Theoretical and applied climatology*, *95*(1-2), 91-109.

[8]. Ghaju, S., & Alfredsen, K. (2012). Evaluation of satellite based precipitations and their applicability for rainfall runoff modelling in Narayani Basin of Nepal. *Journal of Hydrology and Meteorology*, *8*(1), 22-31.

[9]. Habib, E., Haile, A. T., Sazib, N., Zhang, Y., & Rientjes, T. (2014). Effect of bias correction of satellite-rainfall estimates on runoff simulations at the source of the Upper Blue Nile. *Remote Sensing*, *6*(7), 6688-6708.

[10]. Han, D., & Bray, M. (2006). Automated Thiessen polygon generation. *Water resources research*, *42*(11).

[11]. Michaelides, S. C. (Ed.). (2008). *Precipitation: Advances in measurement, estimation and prediction*. Springer Science & Business Media.

[12]. Huffman, G. J., & Bolvin, D. T. (2013). TRMM and other data precipitation data set documentation. *NASA, Greenbelt, USA*, *28*(2.3), 1.

[13]. Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Stocker, E. F., & Tan, J. (2017). V04 IMERG Final Run Release Notes. *NASA Goddard Earth Sciences Data and Information Services Center: Greenbelt, MD, USA*.

[14]. Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., & Stocker, E. F. (2007). The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of hydrometeorology*, *8*(1), 38-55.

[15]. Jiang, S. H., Ren, L. L., Yong, B., Yang, X. L., & Shi, L. (2010). Evaluation of high-resolution satellite precipitation products with surface rain gauge observations from Laohahe Basin in northern China. *Water Science and Engineering*, *3*(4), 405-417.

[16]. Joyce, R. J., Janowiak, J. E., Arkin, P. A., & Xie, P. (2004). CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *Journal of hydrometeorology*, *5*(3), 487-503.

[17]. Kidd, C., & Levizzani, V. (2011). Status of satellite precipitation retrievals. *Hydrology & Earth System Sciences*, *15*(4).

[18]. Kubota, T., Shige, S., Hashizume, H., Aonashi, K., Takahashi, N., Seto, S., ... & Iwanami, K. (2007). Global precipitation map using satellite-borne microwave radiometers by the GSMaP project: Production and validation. *IEEE Transactions on Geoscience and Remote Sensing*, *45*(7), 2259-2275.

[19]. Masood, M., Shakir, A. S., & Rehman, H. U. (2018). Assessment of GPM based Integrated Multi-

satellite Retrievals (IMERG) Under Diverse Climatic Conditions in Pakistan. *Pakistan Journal of Engineering and Applied Sciences*.

[20]. Michaelides, S., Levizzani, V., Anagnostou, E., Bauer, P., Kasparis, T., & Lane, J. E. (2009). Precipitation: Measurement, remote sensing, climatology and modeling. *Atmospheric Research*, *94*(4), 512-533.

[21]. Moazami, S., Golian, S., Kavianpour, M. R., & Hong, Y. (2014). Uncertainty analysis of bias from satellite rainfall estimates using copula method. *Atmospheric Research*, *137*, 145-166.

[22]. Omondi, C. K., Rientjes, T. H. M., Maathuis, B. H. P., & Gumindoga, W. (2017, July). Assessment of bias corrected satellite rainfall products for streamflow simulation: A TOPMODEL application at the headwater catchment of the Zambezi Basin: powerpoint. In *IAHS 2017 Scientific Assembly: Water and Development: scientific challenges in addressing societal issues"*.

[23]. Pancholi, V. H., Lodha, P. P., & Prakash, I. (2015). Estimation of runoff and soil erosion for Vishwamitri river watershed, Western India using RS and GIS. *American Journal of Water Science and Engineering*, *1*(2), 7-14.

[24]. Rehman, A., Jingdong, L., Du, Y., Khatoon, R., Wagan, S. A., & Nisar, S. K. (2016). Flood disaster in Pakistan and its impact on agriculture growth (a review). *Environ Dev Econ*, *6*(23), 39-42.

[25]. Saber, M., & Yilmaz, K. K. (2018). Evaluation and bias correction of satellite-based rainfall estimates for modelling flash floods over the Mediterranean region: application to Karpuz River Basin, Turkey. *Water*, *10*(5), 657.

[26]. Salma, S., Rehman, S., & Shah, M. A. (2012). Rainfall trends in different climate zones of Pakistan. *Pakistan Journal of Meteorology*, *9*(17).

[27]. Sehgal, J. L. (1974). Classification and distribution of the soils of Punjab. In *Proc. Indian Natl. Sci. Acad* (Vol. 40, pp. 404-419).

[28]. Setiawati, M. D., & Miura, F. (2016). Evaluation of GSMaP Daily Rainfall Satellite Data for Flood Monitoring: Case Study—Kyushu Japan. *Journal of Geoscience and Environment Protection*, *4*(12), 101.

[29]. Singh, J., Knapp, H. V., Arnold, J. G., & Demissie, M. (2005). Hydrological modeling of the Iroquois river watershed using HSPF and SWAT 1. *JAWRA Journal of the American Water Resources Association*, *41*(2), 343-360.

[30]. Sorooshian, S., Hsu, K. L., Gao, X., Gupta, H. V., Imam, B., & Braithwaite, D. (2000). Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. *Bulletin of the American Meteorological Society*, *81*(9), 2035-2046.

[31]. Tesfagiorgis, K., Mahani, S. E., Krakauer, N. Y., & Khanbilvardi, R. (2011). Bias correction of satellite rainfall estimates using a radar-gauge product a case study in Oklahoma (USA). *Hydrology & Earth System Sciences*, *15*(8).

[32]. Tian, Y., Huffman, G. J., Adler, R. F., Tang, L., Sapiano, M., Maggioni, V., & Wu, H. (2013). Modeling errors in daily precipitation measurements: Additive or multiplicative?. *Geophysical Research Letters*, *40*(10), 2060-2065.

[33]. Ushio, T., Sasashige, K., Kubota, T., Shige, S., Okamoto, K. I., Aonashi, K., ... & Oki, R. (2009). A Kalman filter approach to the Global Satellite Mapping of Precipitation (GSMaP) from combined passive microwave and infrared radiometric data. *Journal of the Meteorological Society of Japan. Ser. II*, *87*, 137-151.

[34]. Valdés-Pineda, R., Demaría, E., Valdés, J. B., Wi, S., & Serrat-Capdevilla, A. (2016). Bias correction of daily satellite-based rainfall estimates for hydrologic forecasting in the Upper Zambezi, Africa. *Hydrology and Earth System Sciences Discussions*, 1-28.

[35]. Vernimmen, R. R. E., Hooijer, A., Mamenun, N. K., Aldrian, E., & Van Dijk, A. (2012). Evaluation and bias correction of satellite rainfall data for drought monitoring in Indonesia.

[36]. Wardah, T., Bakar, S. A., Bardossy, A., & Maznorizan, M. (2008). Use of geostationary meteorological satellite images in convective rain estimation for flash-flood forecasting. *Journal of Hydrology*, *356*(3-4), 283-298.

[37]. Xie, P., & Arkin, P. A. (1995). An intercomparison of gauge observations and satellite estimates of monthly precipitation. *Journal of Applied Meteorology*, *34*(5), 1143-1160.

[38]. Xie, P., Joyce, R., Wu, S., Yoo, S. H., Yarosh, Y., Sun, F., & Lin, R. (2017). Reprocessed, bias-corrected CMORPH global high-resolution precipitation estimates from 1998. *Journal of Hydrometeorology*, *18*(6), 1617-1641.

How to cite this article: Abdullah, Habib-Ur-Rehman and Mirza, M. S. (2020). Statistical Analysis and Bias Correction of GSMaP Satellite Rainfall Data for Flash Floods Modelling over the Basin of Ravi River, Pakistan. *International Journal on Emerging Technologies, 11*(5): 545–557.