Proceedings of the 5th International Conference on Civil Engineering for Sustainable Development (ICCESD 2020), 7~9 February 2020, KUET, Khulna, Bangladesh (ISBN-978-984-34-8764-3)

A POSITIVE KRIGING APPROACH FOR MISSING RAINFALL ESTIMATION

Nafisa Israt Jaman*[1] and Sajal Kumar Adhikary[2]

[1]*Assistant Engineer, Local Government Engineering Department (LGED), Bagerhat Sadar, Bagerhat, Bangladesh, e-mail: nafisajaman29@gmail.com*
[2]*Professor, Department of Civil Engineering, Khulna University of Engineering & Technology, Khulna-9203, Bangladesh, e-mail: sajal@ce.kuet.ac.bd*

***Corresponding Author**

ABSTRACT

Rainfall data provide fundamental input for various water resources management applications such as design of hydraulic structures, water budget analysis, streamflow estimation, flood frequency analysis and flood forecasting. Hydrologists are often required to estimate areal average rainfall over the catchment and/or point rainfall values at ungauged locations from observed sample measurements at neighbouring locations. Conventionally, stochastic spatial interpolation methods such as kriging are the most commonly used methods for estimating missing point rainfall values at any desired locations based on the available recorded values at neighbouring gauges. However, traditional kriging offers a major weakness because it requires a priori definition of the mathematical function for the variogram model that represents spatial correlations among data points and thus significantly impacts the performance of the methods. The robustness of kriging methods heavily depends on how the variogram model is constructed. Another limitation of traditional kriging is that negative kriging weights are often obtained as a part of the solution for satisfying the requirement of unbiased constraints in the kriging algorithm. It is the variogram that determines the magnitudes of negative weights based on the degree of continuity of the variable. Since positive weights cannot be obtained based on the solutions of kriging algorithms in many cases and thereby positive estimates of desired variables (rainfall in this study) in target locations cannot be ensured in many hydrological applications. In such case, negative weights (when assigned to high rainfall values) lead to negative estimates of rainfall values at the target or base station, which does not make any physical sense. Therefore, in this study, a positive kriging approach is presented where negative kriging weights can be eliminated through a technique called ‘positive kriging’ in the current study. The proposed positive kriging confirms the estimation of positive weights in the traditional ordinary kriging and hence positive estimates in the target or base station. The approach is applied to estimate missing rainfall values at a rain gauge station (Faridpur station in this study) through spatial interpolation using the historical rainfall data from a network of sixteen raingauge stations for a case study area in Bangladesh. The results indicate that Gaussian variogram is identified as the best fitted variogram model and ordinary kriging with the Gaussian variogram model gives the best estimates of the missing rainfall at the base station. This study conclusively proves that the missing rainfall estimation through spatial interpolation by the proposed positive kriging approach could be a viable option to estimate missing rainfall data in the field of hydrology and water resource engineering.

Keywords: *Positive kriging, Variogram model, Base station, Missing rainfall, Raingauge network.*

1. INTRODUCTION

Rainfall is the key climatic variable for most hydrologic analyses for the effective management of water resources systems. However, in practice, missing values frequently occur in rainfall data and the hydrologic analysis is thus hampered by the shortage of consecutive data (De Silva et al., 2007; Simolo et al., 2010). The presence of missing values in the rainfall data in different countries of the world is a common problem for data analysis. Rainfall data may be missing for various reasons such as loss of yearbooks, human errors, wars, fire accidents, occurrence of high floods, occasional interruptions of automatic stations, instrument malfunctions, and network reorganizations etc. (Simolo et al., 2010).

In order to carry out the effective hydrologic analysis, it is essential to estimate the messing value of daily rainfall data. For this purpose, different authors have suggested suitable methods for estimating the missing values for specific countries or regions using several techniques. Because the performance of any method for estimating missing values generally depends on the nature of the missing mechanism, nature of consecutive occurrences of rainfall, nature of neighboring stations, other intrinsic characteristics of the climate variables, etc. (Little and Rubin, 1987).

Conventionally, variance-dependent stochastic interpolation methods, belonging to the general family of kriging, have been widely applied in hydrological sciences for spatial interpolation of hydrologic variables such as rainfall. These methods are based on the principle of minimizing estimation variances at locations where no measurements are available (Adhikary et al., 2016). Kriging in various forms has been used for estimating missing rainfall data at ungauged locations from point measurements available at surrounding stations (Ashraf et al., 1997). Among the various kriging methods, ordinary kriging (OK) remains one of the most preferred stochastic interpolation methods, which has been adopted for estimating missing rainfall values at an ungauged location in a catchment or a region. The performance of OK is highly influenced by the variogram model that represents spatial correlations among data points.

In general, OK does not make sure of getting positive weights and thereby positive estimates of target variable (rainfall in this case) through spatial interpolation. Negative weights can be obtained in OK as a part of the solution for satisfying the requirement of unbiasedness constraints of kriging algorithm (Isaaks and Srivastava, 1989). In case of OK based missing rainfall estimation, negative weights (when assigned to high rainfall values) may lead to the negative estimates of rainfall values at the target or base station, which does not make physical sense. Szidarovszky et al. (1987) and Deutsch (1996) suggest that negative kriging weights should be corrected if it is obtained as a part of the solution. Therefore, a positive kriging approach is presented in the current study where negative kriging weights can be eliminated through a technique called 'positive kriging'. The proposed positive kriging confirms the estimation of positive weights in the traditional OK and hence positive estimates in the target or base station.

2. POSITIVE KRIGING APPROACH

Kriging, the best linear unbiased estimator, in geostatistics refers to a family of generalized least-square regression methods (Isaaks and Srivastava, 1989; Webster and Oliver, 2007). It helps to estimate the unknown variable values at unobserved locations based on the observed known values at surrounding locations. The general expression of ordinary kriging (OK) to estimate missing value of variable *Z* in space is given by:

$$Z_{OK}^{m}(x_0) = \sum_{i=1}^{n} w_i^{OK} Z(x_i) \qquad (1)$$

where, $Z_{OK}^{m}(x_0)$ refers to the estimated missing value of variable Z (rainfall in this study) at desired location x_0; w_i^{OK} is the kriging weights associated with the observation at location x_i with respect to

x_0; and n indicates the number of observed data points. The kriging weights w_i^{OK} mainly depend on the fitted variogram model.

The unbiasedness condition in the kriging estimates is ensured by enforcing a constraint on the kriging weights that is expressed by:

$$\sum_{i=1}^{n} w_i^{OK} = 1 \tag{2}$$

However, the unbiasedness condition indicated in Eq. (2) cannot ensure of getting positive kriging weights in the solution of kriging algorithm. Szidarovszky et al. (1987) suggests that inclusion of an additional non-negative constraint for weights, w_i (i.e., $w_i \geq 0$ where i = 1, 2, 3,…….., n) in the kriging process confirms the estimation of positive kriging weights.

In this study, a variant of positive kriging technique (Teegavarapu, 2007; Adhikary et al., 2016) was adopted to restrict the kriging weights to non-negative values. The objective function was the difference between the observed and estimated rainfall values by the OK method (using the OK-derived weights) over a given time period. The optimization approach used in the proposed positive kriging technique based on mathematical programming formulation can be expressed as:

Minimize

$$\sum_{j=1}^{n}\left[\sum_{i=1}^{N}\left(w_i Z_i^j\right) - Z_m^j\right]^2 \tag{3}$$

Subject to

$$\sum_{i=1}^{N} w_i = 1 \tag{4}$$

$$w_i \geq 0 \tag{5}$$

where Z_m^j is the observed rainfall value at the target or base station (where estimation is desired), Z_i^j is the observed rainfall at individual stations, *j*, *N* is the number of stations excluding the base station, *n* is the number of days (i.e. the specific time period for which data used in individual stations, *j*).

The objective function described by Eq. (3) minimizes the difference between the observed and kriging based estimates of rainfall values over a period of *n* days. The constraint expressed by Eq. (4) makes sure that the estimate is unbiased whereas the additional inequality constraint defined by Eq. (5) will ensure the computation of non-negative kriging weights. The optimization formulation expressed in Eqs. (3) - (5) was solved using the Microsoft Excel Solver with initial weights obtained by the traditional OK method. The solver uses a generalized reduced gradient (GRG) non-linear optimization algorithm for the optimal solution.

The proposed 'positive kriging' approach is applied to estimate missing rainfall values at a rain gauge station (Faridpur station in this study) through spatial interpolation using the historical rainfall data from a network of sixteen raingauge stations for a case study area in Bangladesh, which is shown in Figure 1. As can be seen from the figure, the study area is located in the central part of Bangladesh and there are seventeen (17) rainfall stations operated by Bangladesh Meteorological Department (BMD) enclosed by the large circle. These 17 stations are considered for the analysis and missing rainfall estimation. Details of these rainfall stations are presented in Table 1. Among all these stations, Faridpur station is taken as a base station, where estimation of missing rainfall values will be carried out using the known rainfall values of the remaining sixteen (16) rainfall stations. The base station can be defined as a station where it is assumed that rainfall data are missing but rainfall data are actually available in the location (Adhikary et al., 2016). In this way, the estimated rainfall using kriging technique and the observed rainfall data in the base station can be compared to evaluate the efficacy of the method.

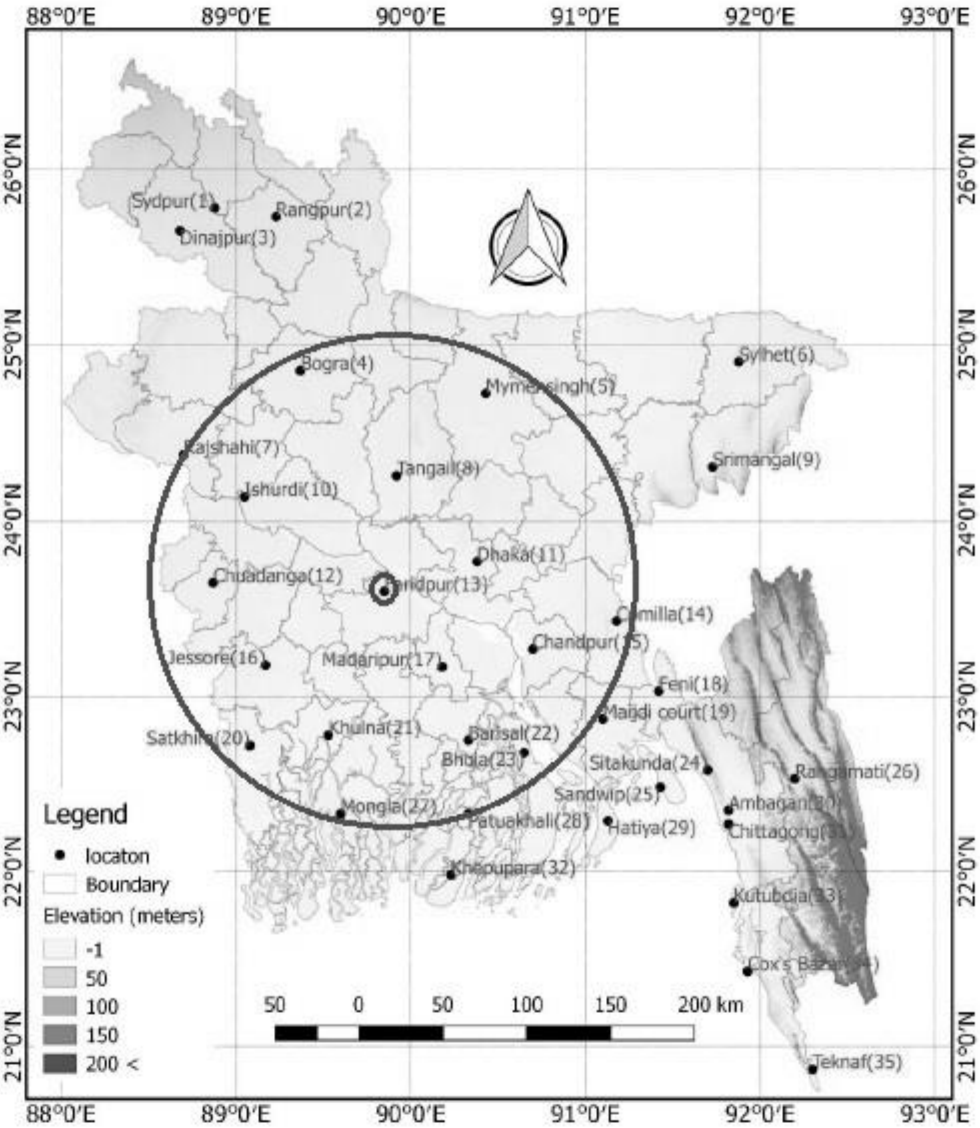

Figure 1: The case study area (enclosed by the large circle) showing the location of rainfall stations (Faridpur is the base station (enclosed by the small circle) at which missing rainfall is to be estimated)

Table 1: Details of raingauge station used in this study

Sl. No.	Station Name	Latitude (Deg)	Longitude (Deg)	UTMX (m)	UTMY (m)
1	Barishal	22.75	90.33	225802.973	2518317.125
2	Bhola	22.68	90.65	258556.565	2510006.493
3	Bogra	24.85	89.37	739490.028	2750420.964
4	Chandpur	23.27	90.70	264720.137	2575275.519
5	Chuadanga	23.65	88.87	690733.691	2616726.444
6	Comilla	23.43	91.18	314058.126	2592296.490
7	Dhaka	23.77	90.38	232982.262	2631224.617
8	**Faridpur***	23.60	89.85	790847.540	2612839.729
9	Ishurdi	24.13	89.05	708327.510	2670143.497
10	Jessore	23.18	89.17	722125.849	2565102.189
11	Khulna	22.78	89.53	759756.056	2521387.708
12	Madaripur	23.17	90.18	211284.241	2565136.011
13	Mongla	22.33	89.60	767814.674	2471663.874
14	Mymensingh	24.72	90.43	240020.327	2736383.727
15	Rajshahi	24.37	88.70	672427.673	2696247.395
16	Satkhira	22.72	89.08	713631.854	2514022.525
17	Tangail	24.25	89.92	796505.415	2685010.465

Note: Faridpur* station is assumed as the base station in the current study, where missing rainfall estimation is to be done.

3. ESTIMATION OF MISSING RAINFALL

3.1 Variogram Modelling

Daily rainfall records from 1980 to 2013 for all seventeen (17) rainfall stations (as shown in Figure 1) are collected from BMD, which are used for the analysis. Summary statistics of collected rainfall data are presented in Table 2. Based on the mean daily rainfall values obtained for all 17 stations, estimation of experimenal variogram is done. Initially, a variogram cloud is carried out and then the variogram cloud is averaged for different lag distances to obtain the experimental variogram.

Table 2: Summary of statistics of collected daily rainfall data

Sl. No.	Station Name	Mean	Std. Dev.	Skewness	Kurtosis
1	Barishal	5.698	15.219	5.112	39.924
2	Bhola	6.208	16.602	5.056	38.674
3	Bogra	4.803	14.330	5.697	49.015
4	Chandpur	5.942	16.790	6.202	65.721
5	Chuadanga	4.057	12.594	6.618	70.911
6	Comilla	5.669	15.768	5.377	47.464
7	Dhaka	5.704	15.809	5.560	54.226
8	**Faridpur**	5.006	14.308	6.098	67.931
9	Ishurdi	4.097	12.050	5.232	39.157
10	Jessore	4.641	13.603	6.190	61.782
11	Khulna	4.963	14.226	7.076	105.753
12	Madaripur	5.396	15.007	5.073	36.572
13	Mongla	5.272	14.071	4.737	31.419
14	Mymensingh	6.196	16.725	5.342	43.926
15	Rajshahi	3.983	12.358	5.971	54.760
16	Satkhira	4.756	13.559	5.859	57.753
17	Tangail	4.950	14.420	5.844	57.526

Finally, the experimental variogram is fitted to a modeled variogram. The most commonly used variogram model included exponential, Gaussian and spherical variogram model functions (Adhikary et al., 2016). They are used to fit the experimental variogram in order to obtain the variogram model. All the best fitted variogram models with the experimental variogram are shown in Figures 2.

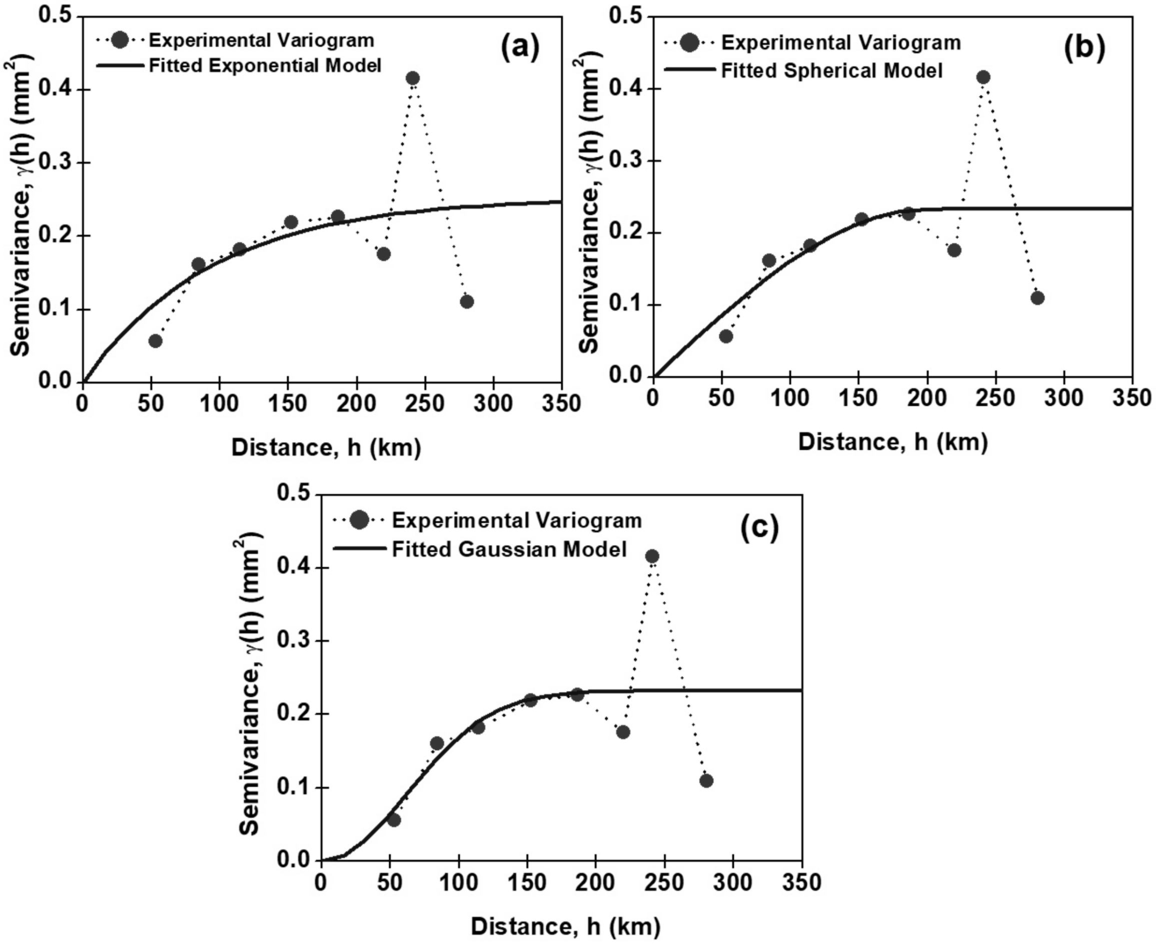

Figure 2: Experimental variogram and best fitted exponential, spherical and Gaussian variogram models for the mean daily rainfall data

While fitting and finding the best fitted variogram models, the corresponding variogram parameters namely, sill, nugget and range are also calculated. Details of the variogram parameters are presented in Table 3. In order to fit the variogram model with the experimental variogram, minimizing the residual sum of squares (RSS) are considered as an objective function. The model which gives the lowest RSS value, is identified as the best fitted model. The variogram modelling and parameters estimation is carried out in the GS+ software platform. As can be seen from Table 3, the Gaussian variogram model gives the lowest RSS value and hence gives the best fitted variogra model. This is also justified from figure 2 and it is seen that the gaussian variogram best fits the first five points of the experimental variogram compared to the remaining variogram models.

Table 3: Summary of variogram parameters of variogram models

Variogram Model	**Nugget (mm^2)**	**Sill (mm^2)**	**Range (Km)**	**RSS**
Exponential variogram	0.0001	0.2532	282.000	0.0562
Spherical variogram	0.0001	0.2332	198.300	0.0536
Gaussian variogram	0.0001	0.2322	150.861	0.0526

Note: RSS = Residual sum of squares

3.2 Estimation of Kriging Weights

After estimating the variogram parameters and identifying all three variogram models including exponential, spherical and Gaussian variogram models, kriging weights are computed by solving the system of simultaneous linear equations in the kriging process. To accomplish this, a spreadsheet application is developed in the Microsoft Excel platform. The estimated kriging weights by different kriging methods including ordinary kriging with exponential variogram model, ordinary kriging with spherical variogram model, and ordinary kriging with Gaussian variogram model are shown in Figures 3.

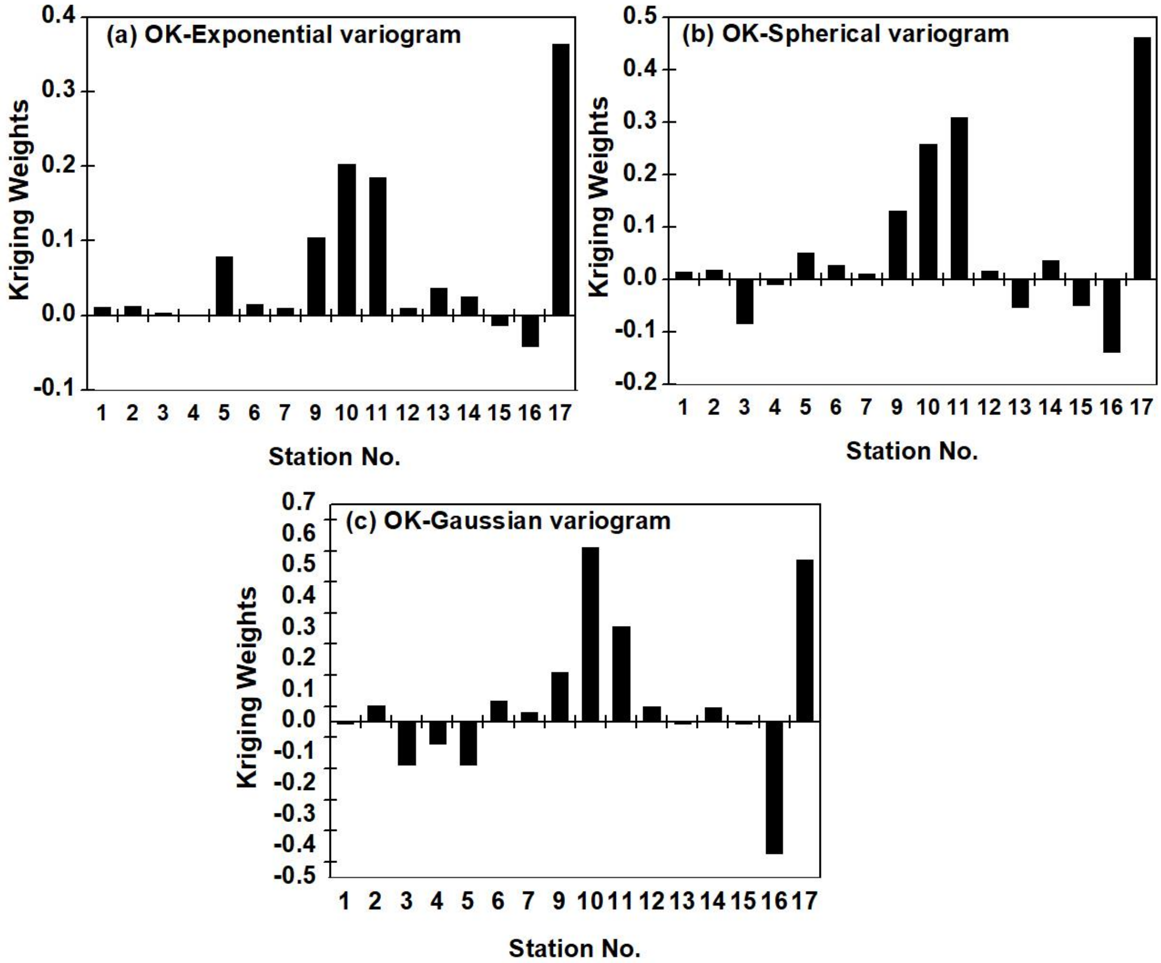

Figure 3: Estimated kriging weights with respect to the base station, rainfall station no. 8 (Faridpur station in this study) using different kriging methods

It is seen from the figure that kriging weights obtained for some of the rainfall stations are negative, which should be made positive before missing rainfall estimation. Since there is no non-negativity constraints in the kriging algorithm, these negative kriging weights are obtained in order to maintain the unbiasedness constraints expressed in Eq. (2). This justifies the development of a new kriging technique, which is referred to as the postive kriging in the current study. In the positive kriging technique detailed in Section 2 and optimization formulation expressed in Eqs. (3) – (5), the inclusion of an additional non-negative constraint for weights, w_i (i.e., $w_i \geq 0$ where i = 1, 2, 3,……., n) in the kriging process confirms the elimination of all negative kriging weights (as shown in Figure 3) and thereby ensures the estimation of positive kriging weights. These positive kriging weights will be used to estimate missing rainfall at the base or target station (Faridpur rainfall station in the current study).

3.3 Estimation of Missing Rainfall

Now, the missing rainfall values at the base station (Faridpur station in this study) are estimated using the positive kriging weights obtained as a solution of the optimization formulation given in Eqs. (3) – (5). The conceptual framework for missing rainfall estimation is presented in Figure 4, where it is assumed that rainfall values are missing at the base station (rainfall station no. 8) although rainfall values are available for that station along with remaining sixteen (16) surrounding rainfall stations. The reason is that in this way the estimated rainfall values by different kriging techniques can be compared with the observed rainfall values to check the efficiency of the different positive kriging techniques in missing rainfall estimation.

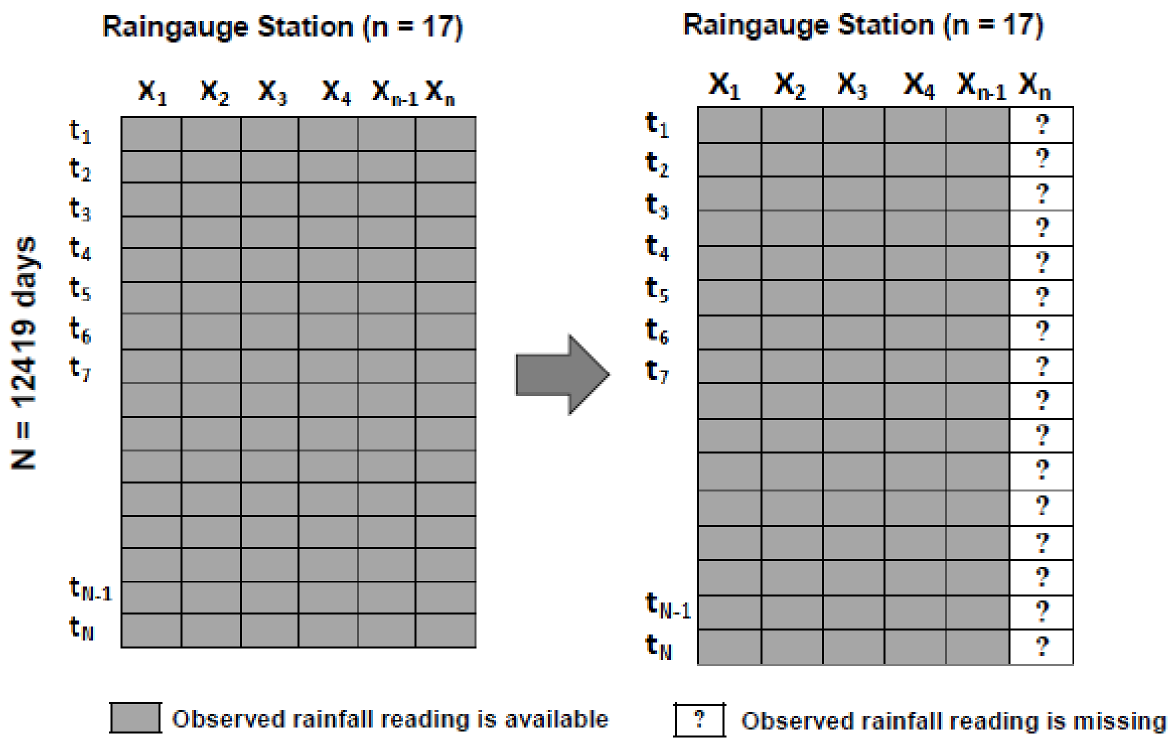

Figure 4: Conceptual framework for missing rainfall estimation at the base station

Now, the observed and estimated rainfall values using different kriging techniques are compared and different error indices are calculated including root mean squared error (RMSE), mean absolute error (MAE) and coefficient of determination (R), which are presented in Table 4. The results presented in Table 4 indicate that ordinary kriging with Gaussian variogram model gives the best estimation with the lowest error and the highest coefficient of determination. Therefore, ordinary kriging with Gaussian variogram model is identified as the best kriging technique for missing rainfall estimation at Faridpur station in this study.

Table 4: Performance of different kriging methods for missing rainfall estimation

Method of Estimation	**RMSE**	**MAE**	**R**
OK-Exponential variogram model	11.747	4.559	0.576
OK-Spherical variogram model	11.775	4.581	0.579
OK-Gaussian variogram model	11.637	4.538	0.591

The estimated and observed rainfall values using different kriging techniques at Faridpur rainfall station is also plotted, which are shown in Figure 5. As can be seen from the figure, a reasonably good agreement between the observed and the estimated rainfall values is obtained. This ultimately proves the efficacy of the kriging technique for estimating missing rainfall values at ungauged locations.

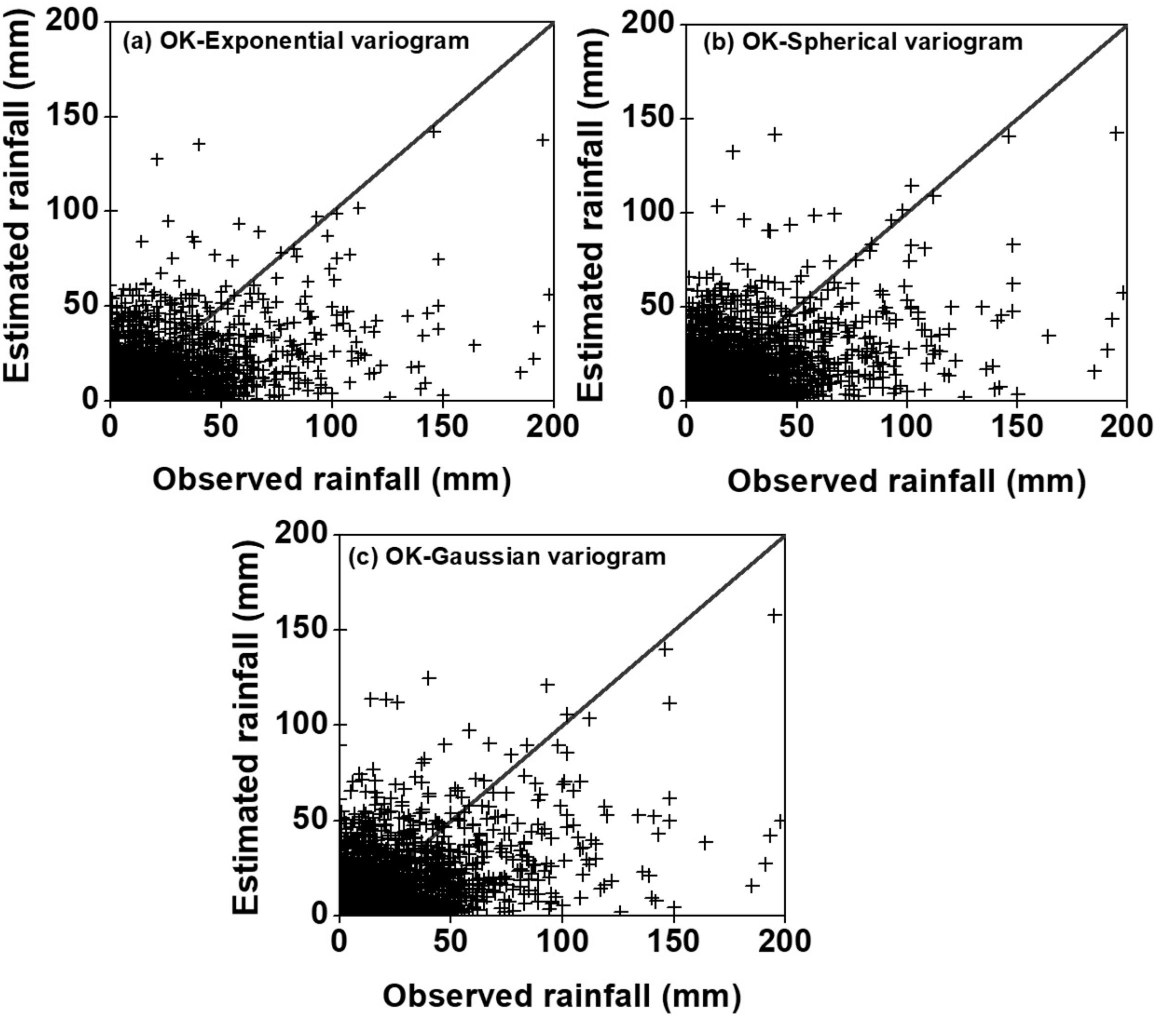

Figure 5: Estimated and observed rainfall values at the base station (Faridpur station in this study) using different kriging methods

4. CONCLUSIONS

In this study, a new variant of kriging called the positive kriging is proposed and adopted to estimate missing rainfall data at the base or target station in a selected area in Bangladesh. Since the traditional kriging does not ensure of getting positive kriging weights, an additional non-negativity constraint has been included and an optimization formulation is developed to achieve the positive kriging weights. The positive kriging weights are used to estimate the missing rainfall values at the base station (Faridpur rainfall station in this study) based on the available rainfall values from sixteen rainfall stations located around the base station. From the experiment, it is obvious that Gaussian variogram is the best fitted variogram model and ordinary kriging (OK) with Gaussian variogram model gives the best estimates of the missing rainfall at the base station. This study conclusively proves that the missing rainfall estimation through spatial interpolation by kriging technique could be a viable option for missing data estimation in the field of hydrology and water resources engineering.

REFERENCES

Adhikary, S.K., Nitin M., & Yilmaz, A.G. (2016). Genetic programming-based ordinary kriging for spatial interpolation of rainfall. *Journal of Hydrologic Engineering*, 21(2), 04015062. DOI: 10.1061/(ASCE)HE.1943-5584.0001300.

Ashraf, M., Loftis. J. C., & Hubard, K. G. (1997). Application of geostatistics to evaluate partial weather station network. *Agricultural and Forest Meteorology*, 84(3-4), 255-271.

De Silva, R.P., Dayawansa, N.D.K., & Ratnasiri, M.D. (2007). A comparison of methods used in estimating missing rainfall data. *Journal of Agricultural Sciences – Sri Lanka*, 3(2), 101–108.

Deutsch, C. V. (1996). Correcting for negative weights in ordinary kriging. *Computers & Geosciences*, 22(7), 765-773.

Isaaks, H. E., & Srivastava, R. M. (1989). *An Introduction to Applied Geostatistics*, Oxford University Press, New York, USA.

Little, J.R.A., & Rubin, D.B. (1987). *Statistical Analysis with Missing Data*. Wiley, New York.

Simolo, C., Brunetti, M., Maugeri, M., & Nanni, T. (2010). Improving estimation of missing values in daily precipitation series by a probability density function-preserving approach. *International Journal of Climatology*, 30: 1564–1576.

Szidarovszky, F., Baafi, E. Y., & Kim, Y. C. (1987). Kriging without negative weights. *Mathematical Geology*, 19(6), 549-559.

Teegavarapu, R. S. V. (2007). Use of universal function approximation in variance-dependent surface interpolation method: an application in hydrology. *Journal of Hydrology*, 332(1-2), 16-29.

Webster, R., & Oliver, M. A. (2007). *Geostatistics for Environmental Scientists* (2nd Ed.), John Wiley & Sons, Chichester, UK.