Comparison of spatial interpolation methods for estimation of weekly rainfall in West Bengal, India

MADHURIMA DAS, ARNAB HAZRA*, ADITI SARKAR*, SABYASACHI BHATTACHARYA*

and PABITRA BANIK*

*Department of Civil Engineering, Jadavpur University, Kolkata, India; *AERU, Indian Statistical Institute, Kolkata, India* (*Received 1 October 2015, Accepted 1 August 2016*) **e mail : banikpabitra@gmail.com**

सार – मौसम की परिघटनाओं में से वर्षा, एक ऐसी घटना है जिस पर अत्यधिक अर्थपूर्ण शोध किए गए जो समकालिक व व्यावहारिक रूप से कृषि संबंधी कार्यों को प्रभावित करते हैं। विशेषकर उन आर्द्र, उप-कटिबंध क्षेत्रों में, जहाँ पर कृषि मुख्य रूप से वर्षा पर निर्भर रहती है। पश्चिम बंगाल के अधिकांश क्षेत्रों में सिंचाई की सुविधाओं की कमी के कारण कृषि उत्पादन के लिए वर्षा एक प्रमुख प्राचल है इसलिए पश्चिम बंगाल मैं वर्षा वितरण के पैटर्न की विस्तृत सूचना अत्यंत महत्वपूर्ण है। वास्तव में, पूरे राज्य में फैले कुछेक विरले स्टेशनों पर ही वर्षा के आँकड़े एकत्रित किए जाते हैं। तथापि वर्षा एक निरंतर होने वाली परिघटना है। अत: तर्कसंगत मानों को सतत स्थानिक पैटर्न में अंतरित करने के लिए संतुलित स्थानिक अंतर्वेशन तकनीक को लागू करना आवश्यक है। इस अध्ययन में अत्यधिक दक्ष अंतर्वेशन तकनीक का पता लगाने के लिए तीन स्थानिक अंतर्वेशन मॉडलों नामत: क्रीगिंग, इनवर्स डिसटेंस वेटिड (IDW) और SPLINE का उपयोग तुलनात्मक विश्लेषण करने के लिए किया गया है। इस विश्लेषण के लिए 19 मानक मौसम विज्ञान सप्ताहों (SMW), 22 से 40 सप्ताहों के लिए 1901 और 1985 के मध्य उपलब्ध वर्षा के साप्ताहिक औसत आँकड़ों का उपयोग किया गया है। तीन अंतवेशन तकनीको की त्रुटियो का विश्लेषण किया गया और न्यूनतम माध्य निरपेक्ष विचलन (MAD) और न्यूनतम विचलन वर्ग माध्य (MSD) मापदंड पर आधारित सर्वोत्तम पद्धति को चुना गया। IDW अंतर्वेशन सर्वोत्तम स्थानिक अंतर्वेशन मॉडल पाया गया है।

ABSTRACT. Rainfall is one of the most eloquently researched contemporary meteorological phenomena affecting the agricultural practices dramatically, particularly along the humid, sub-tropics, where agriculture is predominantly rainfed. It is a key parameter of agricultural production in West Bengal due to lack irrigation facilities in most of the areas. Thus, it is very important to have detailed information of rainfall distribution pattern of West Bengal. In practice rainfall data is collected only at few discrete stations scattered all over the whole state. However, rainfall is a spatially continuous phenomenon rather than discrete. Thus it becomes essential to apply a robust spatial interpolation technique to transform the discrete values into a continuous spatial pattern. In the present study, three spatial interpolation techniques namely Kriging, Inverse Distance Weighted (IDW) and SPLINE, are used for a comparative analysis to identify the most efficient interpolation technique. Weekly average rainfall data available between 1901 and 1985 for 19 standard meteorological weeks (SMW), Week 22 to Week 40 are used for the analysis. The errors of the three interpolation techniques are analyzed and the best method is chosen based on the minimum mean absolute deviation (MAD) and the minimum mean squared deviation (MSD) criteria. The IDW method is found to be the best spatial interpolation technique.

Key words – Agro-meteorology, KRIGING, IDW, SPLINE, MAD, MSD.

1. Introduction

The agricultural system of West Bengal is basically rainfed in nature, *i.e*., dependent on rainfall. Rainfall being a key parameter of climate, the erratic distribution of rainfall becomes a major cause of crop failure. Accordingly, having a brief knowledge of the rainfall distribution patterns is an important aspect to achieve sustainability in agricultural production. In the rainfed ecosystems availability of the rainwater is thus one of the most important considerations of the agricultural planners and policy makers. Therefore, it is essential to gain the knowledge of the rainfall distribution patterns to achieve sustainability in crop production. As it is not possible to observe rainfall values at any spatial location except by a satellite, the data is available only at a few monitoring sites and hence the rainfall data is a spatially discrete realization of a continuous process. However, in reality,

any spatial coordinate has its corresponding rainfall distribution and it cannot be limited by administrative boundaries. The most logical discourse would be the application of Spatial Analysis to interpolate the rainfall observations in order to obtain a smoothened surface.

Several models of spatial interpolation have been applied to climate data such as temperature, precipitation, and wind speed (Phillips *et al*., 1992; Collins and Bolstad, 1996; Goovaerts, 2000; Price *et al*., 2000; Jarvis and Stuart, 2001; Vicente-Serrano *et al*., 2003; Luo *et al*., 2008). The performances of the spatial interpolation techniques are dependent on the geology, and different methods give best fit under different geological conditions. As of author's knowledge, no study till date has explored it in the context of agricultural meteorology of India or particularly West Bengal.

Selecting an appropriate spatial interpolation method is fundamental to surface analysis since different methods of interpolation can result in different surfaces and ultimately different results. Statistical techniques are used to evaluate the three interpolation methods against independently collected data (Legendre, 1998).

 In this study a comparative study is attempted between the three most popular spatial interpolation techniques - KRIGING, Inverse Distance Weighting (IDW), and SPLINE. Estimates are tested against data collected independently, and the benefits and limitations of these methods are discussed. An attempt has been made to analyse if any other model other than KRIGING can yield better interpolation results. Large amounts of time series rainfall data of 79 rainfall stations spread throughout West Bengal and its surrounding area were used to estimate the values of unknown locations. The small scale regional circulations are vulnerable to variations in monsoon rainfall (Rajeevan *et al*., 2008). Thus a general measurement of the strength of monsoon systems is not enough to represent the temporal and spatial distributions. The interpolation techniques are applied to estimate the rainfall distribution pattern of the study area and error from those applied interpolation techniques are evaluated using minimum mean absolute deviation (MAD) and minimum mean squared deviation (MSD). The other criteria generally used for comparison of statistical methods Akaike's Information Criterion (AIC), Akaike's Information Correction Criterion (AICC) and Bayesian Information Criterion (BIC) were not applicable here as the methods SPLINE and IDW are deterministic methods and KRIGING is the only stochastic technique. Therefore, one of the main aims of this study was to apply different interpolation techniques on the rainfall dataset and find out the most suitable one by comparing the values among them.

Fig. 1. Rainfall stations of West Bengal

The paper is organized as follows. In Section 2, materials and methods of the spatial interpolation techniques were discussed. The results of the analysis are provided in Section 3 and Section 4 concludes.

2. Data and methodology

Rainfall data is taken from "Government of India, Weekly Rainfall Probability for Selected Stations of India, India Meteorological Department (IMD), Division of Agricultural Meteorology, Pune". West Bengal boundary was prepared using West Bengal state boundary map. Preparation of the point map of 79 rainfall stations of West Bengal was a basic step to obtain rainfall distribution surface map using interpolation techniques. Locations of 79 rainfall stations were plotted in Fig. 1 according to latitude and longitude information mentioned in the meteorological handbook. Weekly average rainfall data available between 1901 and 1985 for all the 79 stations starting from Week 22 to Week 40 are noted from the handbook. These 19 standard meteorological weeks are considered for the study as they encompass the monsoon season. According to IMD recommendation, the first standard meteorological week refers to $1st$ January to 7th January.

These values are used as input parameter of the spatial interpolation techniques. The Average and interpolated rainfall values for the three selected weeks

are provided in Appendix I, Appendix II and Appendix III.

The three spatial interpolation techniques considered here include two deterministic interpolation methods - Inverse Distance Weighting (IDW) and SPLINE and one stochastic method - KRIGING. All three methods require that the observation locations, also called the monitoring sites, are included in the final output spatial surface. These three spatial interpolation methods have various decision parameters. The selected techniques, KRIGING, IDW, and SPLINE are not the exhaustive set of the interpolation methods used for analyzing spatial data; but are the most common techniques that are available in some GIS software.

(a) *KRIGING* : It is a stochastic technique, and uses a linear combination of weights at known points to estimate the value at an unknown point (Krige, 1966). In contrast with deterministic methods, KRIGING provides a solution to the problem of estimation of the surface by taking account of the spatial correlation. It is a two-step process that begins with semi-variogram estimation and then performs the interpolation. The spatial correlation between the measurement points can be quantified by means of the semi-variogram function:

$$
\gamma(h_1, h_2) = \frac{1}{2N(h_1, h_2)} \sum_{i=1}^{N(h_1, h_2)} \left[R(x_i + h_1, y_i + h_2) - R(x_i, y_i) \right]^2
$$
\n(1)

where, $N(h_1, h_2)$ is the number of pairs of measurement points with distance h_1 along the *X*-axis and *h*2 along the *Y* - axis. Within various variogram models, the spherical model is the most widely used and often preferred when the nugget variance is important and there is a clear range and sill effect (Cressie, 1993; Burrough and McDonnell, 1998). Some advantages of this method are the incorporation of correlation of the variables and the available error surface output. A disadvantage is that it requires substantially more computing and modelling time.

The aim of KRIGING is to estimate the value of an unknown real-valued function, $R(x, y)$ at a point (x, y) given the values of the function at some other points $\{(x_1, y_1), \ldots, (x_N, y_N)\}\)$. A KRIGING estimator is said to be linear because the predicted value $R(x, y)$ is a linear combination that may be written as:

$$
R(x, y) = \sum_{i=1}^{N} \lambda_i R(x_i, y_i)
$$
 (2)

The weights λ_i ; $i = 1, \ldots, N$ are solutions of a system of linear equations which is obtained by assuming that *R* is a sample-path of a spatial Gaussian process $R(x, y)$ and that the error of prediction is to be minimized in some sense.

$$
\varepsilon(x, y) = R(x, y) - \sum_{i=1}^{N} \lambda_i R(x_i, y_i)
$$
 (3)

Here, a nugget term is also considered to estimate the measurement error and the micro-scale variability. ArcGIS 10 is used to prepare spatially interpolated surfaces using the mentioned interpolation techniques.

(b) *Inverse Distance Weighting (IDW)* : It is based on the assumption that the nearby values contribute more to the interpolated values than distant observations. In other words, for this method the influence of a known data point is inversely related to the distance from the unknown location that is being estimated. The advantage of IDW is that it is intuitive and efficient. This interpolation works best with evenly distributed points. Furthermore, unevenly distributed data clusters result in introducing errors. The simplest form of IDW interpolation is called, Shepard method (Shepard, 1968) and it uses weight function w_i given by:

$$
w_i = \frac{d_i^{-p}}{\sum_{j=1}^{N} d_i^{-p}}; i = 1, 2, ..., N
$$
 (4)

where, *p* is an arbitrary positive real number called the power parameter (typically $p = 2$) and d_i are the distances from the dispersion points to the interpolation point, given by:

$$
d_i = \sqrt{(x - x_i)^2 + (y - y_i)^2}
$$
 (5)

 of the unit. Then, the interpolated value of the weekly where (x, y) is the coordinate of the point of interpolation and (x_i, y_i) ; $i = 1, 2,...N$ are the coordinates of the monitoring sites. The weight function varies with a value of unity at the dispersion point to a value close to zero as the distance to the dispersion point increase. The weight functions are normalized as a sum of the weights rainfall $R(x, y)$ is given by:

$$
R(x, y) = \sum_{i=1}^{N} w_i R(x_i, y_i)
$$
 (6)

In order to improve the computation time, it is possible to set bounds to the dispersion points that contribute to the calculation of the interpolated

Fig. 2. Interpolation of weekly rainfall using KRIGING, IDW and SPLINE for 22 SMW

Fig. 3. Interpolation of weekly rainfall using KRIGING, IDW and SPLINE for 28 SMW

Fig. 4. Interpolation of weekly rainfall using KRIGING, IDW and SPLINE for 31 SMW

value, *i.e*., to all those dispersion points within a given search radius centered on the interpolated point. In case the number of monitoring sites, *i.e*., *N* is moderate, it is not necessary to implement this step.

(c) *SPLINE* : SPLINE method estimates values using a mathematical function that minimizes the total surface curvature, resulting in a smooth surface that passes exactly through the sampled points. While there are more entry points specified, the greater the influence of distant points and the smoother the surface. Advantages of splining functions are that they can generate sufficiently accurate surfaces from only a few sampled points and they retain small features. A disadvantage is that they may have

Fig. 5. Interpolation of weekly rainfall using KRIGING, IDW and SPLINE for 34 SMW

Fig. 6. Interpolation of weekly rainfall using KRIGING, IDW and SPLINE for 40 SMW

different minimum and maximum values than the data set and the functions are sensitive to outliers due to the inclusion of the original data values at the sample points. This is true for all exact interpolators, which are commonly used in GIS, but can present more serious problems for SPLINE since it operates best for gently varying surfaces, *i.e*., those having a low variance. Magnus & Clyde (2010) provides a detailed mathematical formulation of the SPLINE interpolation method.

Leave-one-out cross-validation is used to evaluate the performance of each interpolation method. This is achieved by taking one observation out of the sample and predicting it based on the remaining observations at a time. This process allows the mean absolute deviation (MAD) and the mean squared deviation (MSD) test statistics to be calculated for each interpolation method considered in the study.

$$
MAD = \frac{1}{N} \sum_{i=1}^{N} | \hat{R}(x_i, y_i) - R(x_i, y_i) |
$$
 (7)

$$
MSD = \frac{1}{N} \sum_{i=1}^{N} \left[\hat{R}(x_i, y_i) - R(x_i, y_i) \right]^2
$$
 (8)

where, $\hat{R}(x_i, y_i)$ is the predicted value and $R(x_i, y_i)$ the observed value. The MAD and MSD are used to compare different methods by seeing how closely predicted values match the measured values; a technique with smaller values of MAD and MSD is preferred. The performance assessment analysis is done in MATLAB platform (MATLAB, 2012).

3. Results and discussion

Nineteen Standard Meteorological Weeks (SMW) are considered for the undertaken study and for each week, the weekly rainfall is estimated using the three interpolation techniques. The results of each model are presented for five selected weeks as surface maps (Fig. 2 to Fig. 6). The weekly rainfall average magnitude is represented using different shades of blue where the

TABLE 1

MAD values for the three interpolation tecniques for SMW 22 – 40 (mm)

SMW	KRIGING	IDW	SPLINE
22	6.35	3.62	4.29
23	8.73	4.35	5.78
24	9.61	5.63	7.09
25	11.97	7.86	10.64
26	11.02	6.37	7.24
27	12.50	7.09	8.14
28	11.46	6.72	7.88
29	10.24	6.65	8.89
30	10.15	5.69	6.95
31	11.33	6.29	10.35
32	10.57	6.84	8.06
33	9.95	4.60	7.16
34	11.50	5.58	7.45
35	8.56	4.21	5.87
36	9.15	4.96	7.76
37	9.20	4.71	5.00
38	7.52	3.12	5.37
39	7.74	3.92	6.27
40	5.66	3.03	4.72
Seasonal	9.64	5.33	7.10

lower values were represented using lighter shades, and the higher values using darker shade. The interpolated values are valid for the points inside the perimeter of the study site, even when the algorithms that calculate the raster layer, considered all pixels within rectangle defined by the base map of the state boundary.

The results of each model are discussed for the $22nd$, $31st$ and $40th$ week representing the early monsoon, midmonsoon and late monsoon phases over West Bengal. On examining the rainfall distribution for the period of 22 SMW, it is found that the average rainfall was higher in the northern and eastern parts of West Bengal according to kriging interpolation technique. According to IDW techniques, the average rainfall is higher in northern part. The rainfall distribution pattern obtained by SPLINE interpolation however shows a uniform rainfall distribution. On examining the rainfall distribution pattern of 31 SMW, it is found that the average rainfall according to IDW spatial interpolation technique was higher in western and northern parts of West Bengal. The rainfall distribution pattern obtained by SPLINE interpolation however shows a uniform rainfall distribution with an increase from the variation found in the 22 SMW. On examining the rainfall distribution, it is found that the

TABLE 2

MSD values for the three interpolation tecniques for SMW 22 – 40 (mm)

average rainfall of 40 SMW according to Kriging and IDW spatial interpolation techniques are higher in northern and southern parts of the West Bengal. According to SPLINE spatial interpolation technique, it shows an even rainfall distribution pattern. For the rest of the weeks, the trend has remained very much similar.

The performance assessment methods, *e.g*., minimum MAD and minimum MSD criteria yield the results shown in Tables 1&2. They show that among the interpolation methods used, the IDW method is the one that estimates the measurement results of the weekly rainfall the best. The results reflected that the average magnitude of the weekly rainfall is mainly determined by the distance between the sources and the observation point.

Kriging and IDW use a weighting, giving more significance to nearest data points. They both rely on Tobler's first law of geography: things that are close are more related than things that are further apart. As such in IDW points that are far away have far less influence than points that are close. IDW differs from Kriging in that it is simpler and no statistical models are used. It is easy to define and therefore easy to understand the results. In

IDW only known 'z' values and distance weights are used to determine unknown areas. IDW and Spline are sensitive to outliers. Spline is computationally efficient and effective in producing surfaces from regularly spaced data. Unlike IDW, Spline can estimate surface values above as well as below the maximum and minimum values. IDW does not produce values higher than the maximum values, since information of observations decrease with distance from known points.

4. Conclusion

Spatial analysis (spatial interpolation) of climate data (weekly average rainfall) is demonstrated which was useful to predict the unknown quantities of rainfall based on known points. In this study, three spatial interpolation techniques like KRIGING, Inverse Distance Weighted interpolation (IDW), and SPLINE are used to show the rainfall distribution pattern of West Bengal. They all reflect the variable rainfall distribution patterns for West Bengal. A post-interpolation assessment of the performances of each technique is done using MAD and MSD criteria.

In comparison with the other two interpolation methods evaluated, the results of the analysis indicated that IDW is the most likely one to produce the best estimation of a continuous surface for average weekly rainfall and the results have consistency for almost all the 19 Standard Meteorological Weeks, from 22nd to 40th weeks. These weeks are mainly important for agricultural practices during the rainy season. The results are justified in a sense that the correlation of the amount of rainfall decreased with the distance between two points and there are different types of local effects, like the presence of forests, urban areas which regulated the quantity as well. Often the heavy rainfall takes place in the north or south and the other part experienced drought-like situation which clearly illustrates the validity of the results obtained.

KRIGING is undoubtedly the most popular among the three methods considered and often the applied researchers from different disciplines apply this method on their datasets. But as we observed the results, in some situations, the deterministic techniques like IDW yield better results than KRIGING. The other deterministic method SPLINE performs the second best. But it does not conclude that the deterministic methods perform better than stochastic methods. Some non-stationary spatial statistical model would possibly perform better in this situation but any usual software does not take care of a particular type of data. Hence, it is very essential to have a good knowledge about the performance of the commonly used techniques.

Reference

- Burrough, P. A. and McDonnell, R. A., 1998, "Principles of Geographical Information Systems", 1st Edition, Clarendon Press: Oxford.
- Collins, F. C. and Bolstad, P. V., 1996, "A comparison of spatial interpolation techniques in temperature estimation", In Proceedings of the Third International Conference/Workshop on Integrating GIS and Environmental Modelling. National Center for Geographic Information Analysis (NCGIA): Santa Fe, NM, Santa Barbara, CA: January 21-25.
- Cressie, N., 1993, "Statistics for Spatial Data", 1st Edition, John Wiley and Sons: New York.
- Goovaerts, P., 2000, "Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall", *Journal of Hydrology*, **228**, 113-129.
- Jarvis, C. H. and Stuart, N., 2001, "A comparison among strategies for interpolating maximum and minimum daily air temperatures. Part II: the interaction between number of guiding variables and the type of interpolation method", *Journal of Applied Meteorology,* **40**, 1075-1084.
- Krige, D. G., 1966, "Two-dimensional weighted moving average trend surface for ore-evaluation", *Journal of the South African Institute of Mining Metallurgy*, **66**, 13-38.
- Legendre, P. and Legendre, L., 1998, "Numerical Ecology", 2nd Edition, Elsevier, Canada.
- Luo, W., Taylor, M. C. and Parker, S. R., 2008, "A comparison of spatial interpolation methods to estimate continuous wind speed surfaces using irregularly distributed data from England and Wales", *International Journal of Climatology*, **28**, 947-959.
- Magnus, E. and Clyde, M., 2010, "Control Theoretic Splines: Optimal Control, Statistics, and Path Planning", 1st Edition, Princeton University Press.
- Phillips, D. L., Dolph, J. and Marks, D., 1992, "A comparison of geostatistical procedures for spatial analysis of precipitation in mountainous terrain", *Agricultural and Forest Meteorology,* **58**, 119-141.
- Price, D. T., McKenney, D. W., Nalder, I. A., Hutchinson, M. F. and Kesteven, J. T., 2000, "A comparison of two statistical methods for spatial interpolation of Canadian monthly mean climate data", *Agricultural and Forest Meteorology*, **101**, 81-94.
- Rajeevan, M., Gadgil, S., Bhate, J., 2008. "Active and Break Spells of the Indian Summer Monsoon, NCC Research Report: March 2008", Retrieved March 10, 2013. [http://www.imdpune.gov.in/](http://www.imdpune.gov.in/ncc_rept/RESEARCH%20%20REPORT%25) [ncc_rept/RESEARCH%20%20REPORT%](http://www.imdpune.gov.in/ncc_rept/RESEARCH%20%20REPORT%25) 207.pdf.
- Shepard, D., 1968, "A two-dimensional interpolation function for irregularly-spaced data", ACM Annual Conference/Annual Meeting, 517-524.
- Vicente-Serrano, S. M., Saz-Sanchez, M. A. and Cuadrat, J. M., 2003, "Comparative analysis of interpolation methods in the middle Ebro Valley (Spain): Application to annual precipitation and temperature", *Climate Research*, **24**, 161-180.

APPENDIX I

Average and Interpolated Rainfall values for 22 SMW

APPENDIX II

APPENDIX III

Average and Interpolated Rainfall values for 40 SMW

Ξ