



Optimization of Monitoring Network to the Rainfall Distribution by Using Stochastic Search Algorithms: Lesson from Pakistan

TALHA OMER

MAHMOOD UL HASSAN

IJAZ HUSSAIN

MARYAM ILYAS

SYED GHULAM MOHAYUD DIN HASHMI

YOUSAF ALI KHAN

*Author affiliations can be found in the back matter of this article

ORIGINAL RESEARCH
PAPER



STOCKHOLM
UNIVERSITY PRESS

ABSTRACT

Agricultural production is greatly influenced by environmental parameters such as temperature, rainfall, humidity, and wind speed. The accurate information about environmental parameters plays a vital and useful role when making policies for the agriculture sector as well as for other sectors. Pakistan meteorological department observed these environmental parameters at more than 90 stations. The allocation of these monitoring stations is not made systematically correct. This leads to inaccurate predictions for unobserved locations. The study aims to propose a monitoring network by which these prediction errors of the environmental parameters can be minimized. The well-known prediction techniques named, model-based ordinary kriging and model-based universal kriging (UK) with the known Matheron variogram model are used for prediction purposes. We investigate the monitoring network of Pakistan for rainfall and focus on both the optimal deletion/addition of monitoring stations from/to this network. The two stochastic search algorithms, spatial simulated annealing, and genetic algorithm are used for optimization purposes. Furthermore, the minimization of the Average Kriging Variance (AKV) is taken as the interpolation accuracy measure. The spatial simulated annealing exhibits a lower AKV as compared to the Genetic algorithm when adding/removing the optimal/redundant locations from the monitoring network.

CORRESPONDING AUTHOR:

MAHMOOD UL HASSAN

Department of Statistics,
Stockholm University, Sweden
mahmood.ul-hassan@stat.su.se

KEYWORDS:

Environmental parameters;
Variogram; Genetic algorithms;
Spatial simulated annealing;
Average Kriging Variance

TO CITE THIS ARTICLE:

Omer, T, Ul Hassan, M, Hussain, I, Ilyas, M, Din Hashmi, SGM and Khan, YA. 2022. Optimization of Monitoring Network to the Rainfall Distribution by Using Stochastic Search Algorithms: Lesson from Pakistan. *Tellus A: Dynamic Meteorology and Oceanography*, 74(2022), 333–345. DOI: <https://doi.org/10.16993/tellusa.247>

1. INTRODUCTION

Rainfall is one of the major and important environmental parameters which is responsible for floods, droughts, and drastic events like land sliding. The drastic floods of 2009 and 2010 caused more than 43 billion USD to Pakistan's total economy and more than 178 casualties (see, https://en.wikipedia.org/wiki/2010_Pakistan_floods). However, Kirsch et al. (2010), claimed the 1700 casualties and 9.743\$ billion known loss to local infrastructure, which includes school, healthcare units and some other government institutes. Therefore, there is a great need to predict rainfall variables more precisely. Pakistan monitoring stations are not equally distributed across the country. In the north of Pakistan, episodes of heavy rainfall are prevalent. However, there are very few monitoring stations in this area. On the other hand, a large number of monitoring stations are deployed in the southern region fewer episodes of rainfall are recorded there.

There is an impulse need to revisit the distribution of the monitoring network of the rainfall. This can potentially be done by removing redundant locations from the existing monitoring network. Further, new monitoring stations can be deployed to predict the environmental parameters more precisely. Adding optimal and deleting the redundant locations can reduce prediction errors.

Modelling the spatial behavior of different environmental parameters has been studied by many researchers (Hengl et al. 2004, 2007; Ikehukwu et al. 2017; and Yang, 2018). Geostatistical models provide flexible approaches to capture the spatial behavior of the environmental parameters. Spatial prediction or Kriging techniques belongs to a family of geostatistical techniques, that mainly focus to investigate the spatial behavior of the data. Kriging procedures have been widely used for spatial prediction of the environmental parameters, groundwater, droughts, and soil sample prediction at different locations. Zahid et al. (2016) modeled the spatial distribution of the sodium concentration by means of Universal Kriging and Bayesian Universal Kriging in groundwater. It was found that Bayesian Universal Kriging better-fitted the data.

Spatial prediction of sulfate concentration in groundwater has been carried out for the Southern Punjab of Pakistan (Mubarak et al. 2015). A study on mapping and spatial prediction of the average annual precipitation of Turkey by using the five different kriging methods has been carried out by Bostan et al. (2012). The rainfall on 225 meteorological stations was measured and the superiority of universal kriging was showcased. Many simple and complex interpolation methods have been developed to estimate the value of spatially distributed data including the environmental parameters at unobserved locations on the basis of observed location

(see the massive literature e.g. Knotters et al. 1995; Phillips et al. 1997; Carrera-Hernández and Gaskin, 2007; Wang et al. 2002; Wang et al. 2010; Hussain et al. 2014; Omer et al. 2019; and Ellahi et al. 2021).

Modelling the spatial distribution and characteristics of the particular matter remains a focal point of research. There are numerous optimisation methods to identify the optimal solution for any monitoring network. Spatial sampling design is widely used to identify the optimal solution for the monitoring network (see the recent massive literature, e.g., Fuentes et al. 2007; Zhu and Stein, 2006; Zimmerman 2006, Spöck and Pilz, 2010; Hussain et al. 2010; Hussain et al. 2011; Hussain et al. 2015 and Khan et al. 2021). The optimal spatial sampling designs can provide minimum mean square prediction error. The resultant optimized monitoring networks can save cost of installation and manpower. Despite being widely used, spatial sampling design is one of the topics that could not get enough attention in the literature due to laborious and complex mathematical computations. Unfortunately, sometimes the spatial sampling design failed to produce reliable results. In such cases, stochastic search algorithms (Guedes et al. 2011) are used. The stochastic search algorithm has been extensively used in spatial statistics to find the optimal solution for the different monitoring networks (see e.g., Journel, 1990; Deutsch and Cockerham, 1994; Gringarten and Deutsch, 1999; Al-Mudhafar, 2019). The choice of stochastic search algorithms is typically based on how the algorithm deal with arbitrary systems, statistical guarantees for finding an optimal solution, computational cost, and accuracy. The Spatial Simulated Annealing (SSA) has many features of a good search algorithm. Similarly, a Genetic Algorithm (GA) can end up with multiple local optima. It is a global optimization method that can provide a good solution by utilizing the minimum available information. Also, it provides a good and fast convergence towards an optimal solution (Gallagher and Sambridge, 1994). Gallagher and Sambridge (1994) discussed two stochastic search algorithms (GA and SSA). SSA and GA has widely been used to optimize the monitoring network of the rainfall distribution. Pardo-Igúzquiza (1998) established an optimal network design to estimate the areal averages of rainfall events by using the SSA. Recently, Wadoux et al. (2017) utilized the SSA to minimize the space-time average kriging external drift variable to find an optimal solution for the monitoring network of the rainfall network in the north-east of the city of Manchester in the United Kingdom. Adib and Moslemzadah (2016) presented the optimal selection of the rainfall gauging stations by combining the kriging and genetic algorithm methods. The estimation error of different rainfall gauging estimations was calculated using inside and outside stations of the watershed. Each combination

is further given to GA to select the optimal location by minimizing the error variance. More latest literature on SSA and GA to optimize the rainfall monitoring network can be visited (e.g. Nasseri et al. 2008; Soroush & Abedini, 2019; Molla et al. 2022). The next paragraph explains some of the transition histories of GA and SSA in different ways.

The GA's and genetic programming approaches have been the most widely used in credit scoring applications. Goldberg (1975) discussed that GA provides a method to perform a randomized global search in a solution space. Metropolis et al. (1953) introduced a stochastic search relaxation method that can simulate the performance of a system of particles approaching thermal equilibrium. The algorithm compares the energy with some given criteria of different particles to that of a dissimilar algorithm. If the energy of the new configuration is smaller as compared to the previous one, the new formulation is accepted. Fabian (1997) studied the performance of simulated annealing procedures for searching a global minimum of a function. Bélisle (1992) discussed the convergence properties of simulated annealing procedures for continuous functions. These results were applied to hit-and-run algorithms used in global optimization. Fleischer (1996) introduced cybernetic optimization by simulated annealing as a procedure of parallel processing that reduced the processing time for the convergence of simulated annealing to the global optima. Furthermore, Fleischer (1999) extended the theory of cybernetic optimization by simulated annealing into the continuous domain by applying probabilistic feedback control to the generation of candidate solutions. Heuvelink et al. (2010) used the plume simulation to optimize some additional cities by minimizing the predictable cost of a wrong decision and area of false positive and false negative detection. However, this method is very time-consuming, due to geo-statistical simulation which is based on the iterative numerical optimization algorithm.

In this article, we applied the SSA and GA following the ethos of Baume et al. (2011) and Santacruz et al. (2014). Here, the Average Kriging Variance (AKV) for Pakistan Monitoring network to the rainfall distribution is minimized. Two interpolation techniques named model based OK and model based UK are being used to observe the unknown location based on known locations. The novel contribution of this article is that we modified the SSA and GA by using the model based OK and the model based UK to optimize the monitoring network of rainfall distribution in Pakistan.

This paper is distributed in six sections in total. Section 2 represents the detailed description of the data set that has been used for this research. SSA optimization technique with the model based OK and the UK is described in Section 3. Section 4 briefly explains the

GA optimization technique. The results and discussion are provided in Section 5. In the end, some concluding remarks have been given in section 6.

2. PAKISTAN RAINFALL DATA SET

Pakistan is situated between 23°–37° towards north latitude and 61°–76° towards east longitude. Pakistan holds the deserts in western areas. These areas are observed to have high temperatures and remain dry. The warm season and little precipitation are observed throughout the year in the coastal areas which are situated along the Arabian Sea. The northern areas of Pakistan have mountains. The well-known Karakorum region is also in the north of Pakistan, having the world's largest mountains. These areas are generally very cold with frequent episodes of heavy rainfall. The area between the latitude 24°–30° and longitude 62°–67° is the east-south region of Pakistan. The east-south regions have deserts. This region has low height (less than 150 m) and remains quite hot even in the monsoon period. Pakistan capital Islamabad is located between latitude 32°–35° and longitude 68°–72°. There is usually very heavy rainfall throughout the year in the capital of Pakistan. The areas with heavy rainfall are above thousand-meter on average. The average elevation is 1400 m in the northern areas of Pakistan and these areas remain cold compared to all other regions of Pakistan.

In general, summer season is wet, however the winter remains dry. The average rainfall varies from 200 mm in the north to 30 mm in south.

Pakistan meteorological department is observing environmental variables at more than 90 monitoring stations. The rainfall data are collected from 52 stations from the Pakistan Meteorological Department, Islamabad, from 1998 to 2019. The aim is to cover most of the country, for which fifty-two meteorological stations are sufficient to obtain optimum results. Rainfall is considered as the most responsible environmental parameter for climate change. Pakistan's climate is diverse. Rainfall contributes to natural disasters like floods and land sliding issues. Therefore, modelling the variation in rainfall can help in future planning. For computational ease, the monthly average rainfall is used as the response variable. We use two different optimization techniques. These were used to optimally add and delete 5, 10, 15, 20, 25 and 30 number of locations selected registered cities and from existing monitoring network, respectively.

The study area and existing monitoring network is displayed in Figure 1 above with the layers of average rainfall. On the other hand, Figure 2 shows the study area map consisting of 441 registered cities in Pakistan. from where optimal locations can be chosen from these.

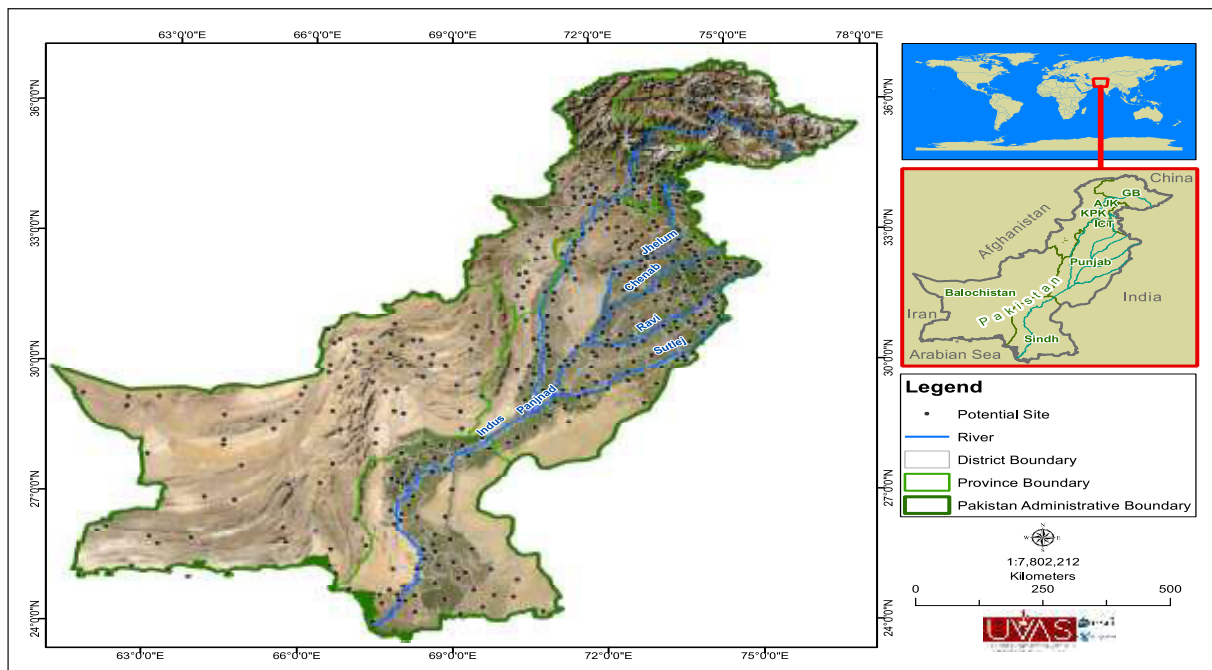


Figure 1 Study area with 441 potential candidates.

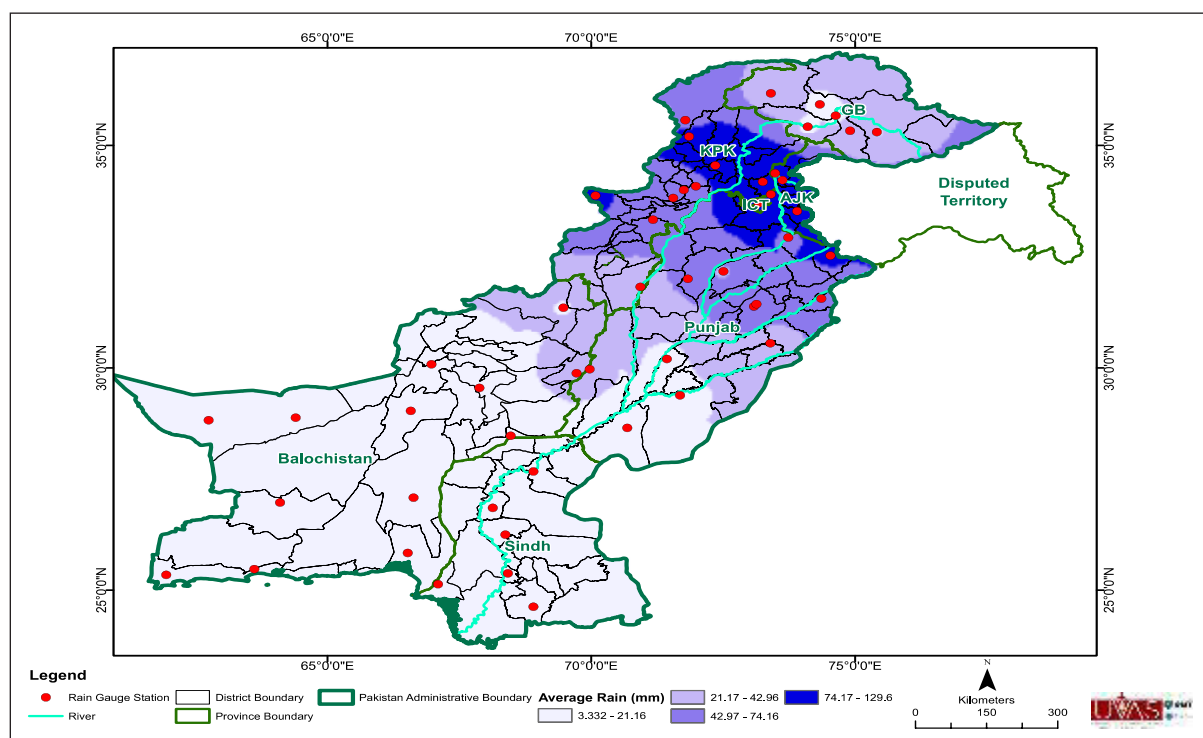


Figure 2 The 52 sampling locations in the existing rainfall monitoring network of Pakistan.

3. SPATIAL SIMULATED ANNEALING (SSA)

The core idea of SSA is to select new locations based on some performance criteria. We modified the SSA by minimizing AKV of the model based on OK and UK, following the ethos of Riberio and Diggle (2007). It starts with the initial design and computation of the objective function AKV. Then, it produces a new design by moving one randomly chosen point from some random direction. Again, compute the objective function for the new design.

Accept the new design with some probability if the objective function is improved. Decrease the probability by increasing the iterations of algorithms. Repeat these steps until the objective is met. It was noted that accuracy increased as we increased the number of iterations of the process. In this study, 10,000 iterations were carried out for achieving the objective. Spatial Simulated Annealing needs some other parameters to be defined as well. For instance, the probability of accepting and rejecting the new design, 'Colling' schedule and a stopping criterion are also required for achieving the underlying objective.

The two well-known interpolation procedures OK and UK with known variogram models are used for simulated annealing. The next two subsections give the detailed description of the used interpolation procedures.

3.1. MODEL BASED ORDINARY KRIGING

Let $Z(s_1), Z(s_2), Z(s_3), \dots, Z(s_n)$ be random variables at spatial locations $s_1, s_2, s_3, \dots$. Kriging is an interpolation technique used to predict the unknown values of a random variable, Z , at one or more unobserved locations, stating differently, kriging is used to interpolate random field, Z , at unobserved locations (Matheron, 1963). The South African engineer D.G. Krige is the pioneer founder of kriging and this technique was named on his name. The main developments came in kriging from G. Matheron after 1960. So far numerous kriging techniques have been developed like Ordinary kriging (OK), Universal Kriging (UK), Simple Kriging (SK), Regression Kriging (RK), Bayesian Kriging (BK), and many more can be found in the literature. One of the major benefits of this technique is that it calculates the prediction error of the response variable. Kriging considers the spatial dependence for mapping purposes, and also for sampling locations. The spatial dependence is further modelled by fitting theoretical variogram model to empirical variogram. The variogram model can be written as

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i,j \in N(h)} \{Z(s_i) - Z(s_j)\}^2 \quad (1)$$

Where $Z(s_i)$ and $Z(s_j)$ are the realizations of the response variable at known locations s_i and s_j respectively, and $N(h)$ is the number of point pairs falling within the distance h (Riberio and Diggle, 2007). OK is one of the efficient techniques which has been used frequently. The basic assumption of this technique is that the mean is considered constant but unknown. This technique has the ability to produce the predicted values and the corresponding prediction errors. This method is believed to consider one of the simple and popular methods in spatial prediction. The response variable Z follows the Gaussian distribution in the model (see details, Riberio and Diggle, 2007). Further details could be found in

$$Z \sim N(\mu, \Sigma_z) \quad (2)$$

where μ is the mean and Σ_z is the covariance matrix and can be written as

$$\Sigma_z = \sigma^2 R(\alpha) + \tau^2 I \quad (3)$$

where R is a correlation matrix and it depends on vector-valued parameter α , ($\sigma^2 = \text{sill}$; $\varphi = \text{range}$, and $\tau^2 = \text{nugget}$). Generally, Kriging techniques require the data to be normally distributed. The parameters of model can be estimated by using well-known estimation methods e.g., Restricted Maximum Likelihood (REML), Weighted Least Square (WLS), Ordinary Least Square (OLS), and Maximum Likelihood (ML).

The variogram model which produces minimum Mean Square Prediction Error (MSPE) is preferred for interpolation of the response variable. The response variable at the unobserved sites can be predicted by the system of ordinary kriging as follows.

$$\hat{Z}_0(s_0) = \sum_{i=1}^N w_i(s_0) Z(s_i)$$

where w_i are the weights to fulfil the condition of unbiasedness.

$$\sum_{i=1}^n w_i(s_0) = 1$$

The Lagrange multiplier can be used to minimize the variation of prediction subject to unbiasedness.

3.2. MODEL BASED UNIVERSAL KRIGING (UK)

A process that offers an estimator including the local trend in the data set is known as Universal Kriging (Isaaks and Srivastava, 1989). In the previous subsection, it is illustrated that in OK, mean remain constant but unknown in whole study area, however in UK, the mean is the function of some local trend's coordinates. Mathematically model of UK can be written as follows

$$Z \sim N(\mu(s), \Sigma_z) \quad (4)$$

where $\mu(s)$ is considered as mean and Σ_z is known as covariance matrix.

$$\Sigma_z = \sigma^2 R(\alpha) + \tau^2 I \quad (5)$$

where R is a correlation matrix and it depends on vector-valued parameter α , ($\sigma^2 = \text{sill}$; $\varphi = \text{range}$, and $\tau^2 = \text{nugget}$). The parameter estimation methods are same which have been used in OK method. The trend component in UK can be modelled as

$$\mu(s) = \sum_{k=1}^l b_k p_k(s), \quad (6)$$

where b_k is the k_{th} coefficient, p_k is the function that defines the trend, and l is the number of functions that is used to model the trend. The remaining process for estimating the weights is similar as given in subsection (3.1).

3.3. CROSS-VALIDATION STATISTICS

Cross-validation statistics are used to compare the performance of kriging methods (Mubarak et al. 2015). Three different methods are frequently used in the literature. These methods are named, Mean Bias Error (MBE), Mean Absolute Error (AE), and MSPE and that are as follows:

$$MBE = \frac{1}{n} \sum_{i=1}^n \{\hat{z}(x_i) - z(x_i)\}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{z}(x_i) - z(x_i)|$$

$$MSPE = \frac{1}{n} \sum_{i=1}^n \{\hat{z}(x_i) - z(x_i)\}^2$$

where $\hat{z}(x_i)$ is the estimated value of data, $z(x_i)$ is the observed value of the data. Here, MSPE is used as criterion to compare the performances of both used interpolation techniques.

4. GENETIC ALGORITHM (GA)

Genetic algorithm (GA) deals population of string which demonstrates the parametrization of the optimization problem. In biological sense, the strings are known as genotype or chromosomes. The actual situation is to map the representation of phenotype. GA includes the characteristic to encourage/discourage the overachieving /under achieving of the string in the population. In this article, we use GA to minimize the AKV as the interpolation accuracy measure. Interpolation procedures are already discussed in section 3.1 and section 3.2. The basic algorithm of GA at first step is to determine and define an initial population, $x_0^{(0)}, x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)}$ and set $k = 0$. Second step, calculate the objective function that is called “fitness” function for every fellow of the population, $f(x_i^{(k)})$ and allocate probabilities p_i to each item in the population, feasibly proportional to its fitness. Third step, select a probability sample of size $m \leq n$. This is the reproducing population. Fourth step, randomly from a new population $x_1^{(k+1)}, x_2^{(k+1)}, x_3^{(k+1)}, \dots, x_n^{(k+1)}$ from the reproducing population, using numerous mutation and recombination rules. This can be done using random sample selection of the rule for each single pair of individuals. Fifth step, if convergence criteria are meet, stop, and convey $\text{argmin}_i x_i^{(k+1)}$ as the optimum; otherwise, set $k = k + 1$ and repeat the procedure from step one.

Few of the main advantages of using the GA is, the GA has ability to solve every optimization problem, and it can solve problems with multiple solutions. GA techniques doesn’t demand the complex mathematical knowledge, and is very easy to implement and understand. The GA is used to point out optimum sites out of the n locations. Pakistan Monitoring rainfall data set of 52 network locations with their respective coordinates is used for the optimization purposes. We add and delete 5, 10, 15, 20, 25 and 30 number of new locations from existing and 441 registered cities of Pakistan, respectively, by minimizing the AKV. The AKV is calculated as

$$AKV = \frac{\sum_i^n \sigma_i^2}{n} \quad (7)$$

where σ_i^2 is the Kriging variance and n is the number of samples. For evaluation of the GA, the population size is defined as 40 individuals, which means that for every n point in the network, there is 40 pairs of coordinates for

each iteration. We take 10000 iterations for addition of our required points that are mentioned above.

5. RESULTS AND DISCUSSION

This section demonstrates the twofold benefits of the of stochastic search algorithms. First, we can save the sampling locations for precise prediction; second, the accurate prediction of the rainfall parameter can be attained with fewer sampling location. In more details, visualization of 52 monitoring station from the original rainfall monitoring network is shown in [Figure 1](#). In [Figure 2](#), we visualize the study area along with additional 441 potential candidates from where we can add the new locations which may be used for improving the original monitoring network of rainfall. We have denser original rainfall monitoring network in the north compared to south. Therefore, the first target is to eliminate the redundant locations from the north of network. In the 2nd step, we add these locations in into the reduced network in an optimal way so that the new design should be space filling and better covering location in the south as well. We used the minimization of AKV given in equation (5) as design criteria for selecting the new locations from the 441 potential candidates. The spatial variability of rainfall is modeled between known and unknown location by Matheron variogram model ([Matheron, 1989](#)). After fitting the variogram model, we found the parameters initial range and spatial sill 8 and 2, respectively. We can consider different techniques to estimate the parameters of Variogram model, but we stick to maximum likelihood (ML) and utilize it to estimate the parameters of variogram model. The estimated nugget and partial sill are 0.2632 and 2.5324, respectively. After estimation of Variogram models parameters, we can say that it shows strong dependencies among the known and unknown locations align with the similar results [Zahid et al. \(2016\)](#). At second step, we made a spatial prediction of rainfall by using the model-based OK and model-based UK on 441 unobserved locations. Mean square prediction error (MSPE) is used as cross validation criteria. Observed MSPE of OK and UK is 39.64 and 37.88, respectively. The MSPE of UK is slightly lower as compared to the OK, which indicates the superiority of UK over OK.

After first two steps, we are ready to design an optimal monitoring network of rainfall distribution. We use the SSA and GA for minimizing the AKV to remove the redundant (5, 10, 15, 20, 25 and 30) locations from the existing monitoring network. We notice that most of the redundant location exists in north side as the monitoring network is denser in north (may be locations are nearer to each other). We predicted the MSPE on the reduced monitoring network and we notice a slight increase in MSPE of OK and UK, from 39.64 to 41.37, 37.88 to 39.34, respectively, while predicting the unobserved locations.

Our results are pretty much aligned with the results of (Spöck and Hussain, 2012) spatial sampling design for Pakistan. We produce the visualization reduced design by removing the 30 redundant locations from existing monitoring network in Figures 3 and 4. We have the name and coordinates of the reduced design with their visualizations (can be provided on the demand). At second

step, we added the (5, 10, 15, 20, 25 and 30) location in existing network of rainfall and reduced monitoring network. These new added/deleted location with SSA and GA are shown in Figures 3 and 4. Now we have multiple reduced optimal networks so we can make a spatial prediction on the reduced optimal network to check the optimality of our method. We have reduced design with

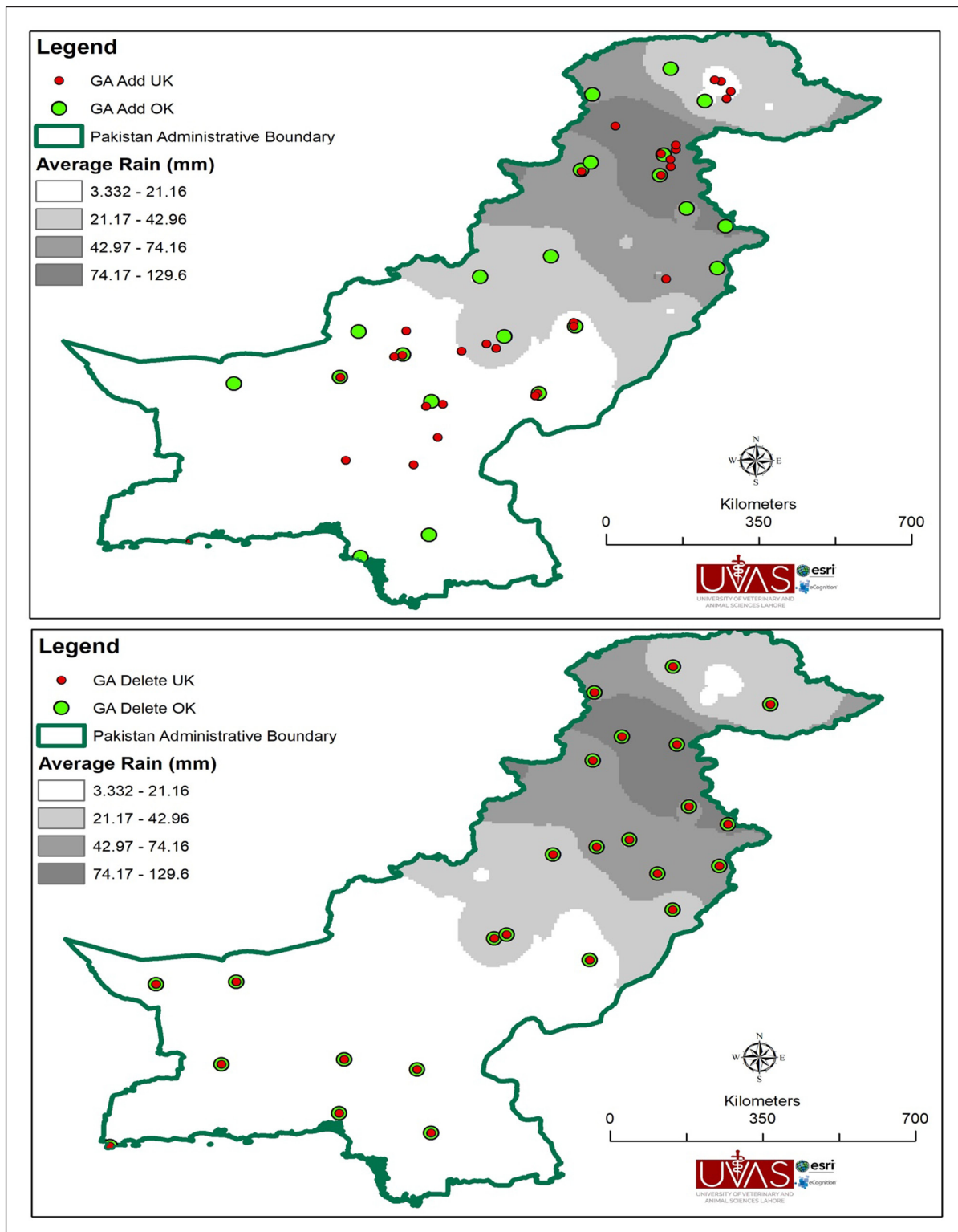


Figure 3 Geographical placement of 30 additional and deleted location by GA under both interpolation procedures for Pakistan meteorological rainfall monitoring network.

Note: Some of the points overlap each other (all the coordinate's names can be provided on demand).

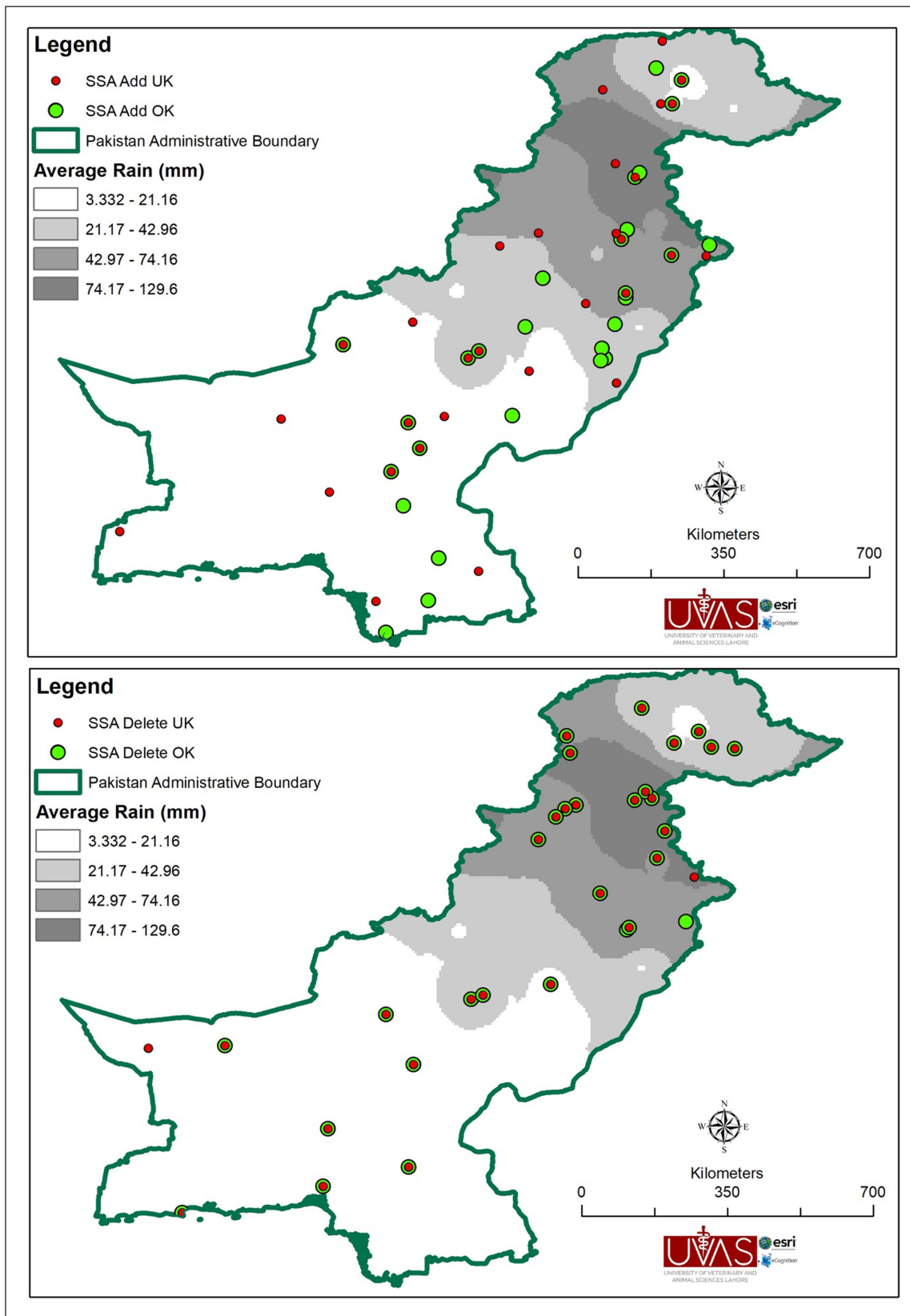


Figure 4 Geographical placement of 30 additional and deleted location by SSA under both interpolation procedures for Pakistan meteorological rainfall monitoring network.

Note: Some of the points overlap each other (all the coordinate's names can be provided on demand).

22 locations (after removing the 30 redundant location) and we add the 30 new locations those are optimally selected by the SSA and GA. We made the prediction of rainfall on 441 unobserved locations with UK and OK

for the optimal added points of SSA and GA. We find a significant decrease in MSPE of UK and OK, in both cases either we add the points by SSA and GA. The MSPE of the OK and UK are recorded to 34.39 and 32.56, respectively,

by considering the SSA optimal points. Similarly, The MSPE of the OK and UK are recorded 36.01 and 33.92, respectively, by considering the GA optimal points. At this stage, we can add some more remarks on the resulting optimal monitoring network of rainfall.

- At first the optimal points are selected in the areas where it looks unfilled, for instance, in our case it looks empty in south. This is obvious that the kriging variance is generally more increased there. This remark is consistent with the finding of Spöck and Hussain (2012).
- As whole we can say that the resulting optimal monitoring network of the rainfall distribution seems to be space filling network: It is a property statement that the examined optimal monitoring network criterion is usually based on accurate prediction not how the covariance function is best estimated. The best estimation of the covariance function also

requires that the new potential candidates' sites should be very near to each other, for instance, proper estimation of the nugget effect.

- Every chosen point selected only once: Our optimal selection of the candidates is based on the without replacement selection of the potential candidates. So, our optimal selection of the points fulfils the criteria, which is usually tough to attain without stochastic search algorithms.

Furthermore, Table 1 represents the AKV by using the SSA and GA to add/delete the specific number of location from/to the existing monitoring network. From the results, it can be observed that SSA remained efficient in all scenarios. We noticed that SSA and GA has less AKV with UK as prediction technique which is pretty much sure and this result is aligned to our presented cross validation results.

SCENARIO	METHOD	IP	AKV	SCENARIO	METHOD	IP	AKV
Add5	SSA	OK	1.1498	Del 5	SSA	OK	1.1935
Add5	SSA	UK	0.8419	Del 5	SSA	UK	0.9723
Add5	GA	OK	5.0918	Del 5	GA	OK	5.6001
Add5	GA	UK	4.8221	Del 5	GA	UK	5.4211
Add10	SSA	OK	1.1181	Del 10	SSA	OK	1.2070
Add10	SSA	UK	0.8399	Del 10	SSA	UK	0.9812
Add10	GA	OK	4.9211	Del 10	GA	OK	5.6821
Add10	GA	UK	4.7811	Del 10	GA	UK	5.5211
Add15	SSA	OK	1.1181	Del 15	SSA	OK	1.2286
Add15	SSA	UK	0.8382	Del 15	SSA	UK	0.9935
Add15	GA	OK	4.8221	Del 15	GA	OK	5.7211
Add15	GA	UK	4.5211	Del 15	GA	UK	5.6244
Add20	SSA	OK	1.0943	Del 20	SSA	OK	1.2563
Add20	SSA	UK	0.8322	Del 20	SSA	UK	1.0001
Add20	GA	OK	4.7221	Del 20	GA	OK	5.8001
Add20	GA	UK	4.4321	Del 20	GA	UK	5.6901
Add25	SSA	OK	1.0832	Del 25	SSA	OK	1.2955
Add25	SSA	UK	0.8298	Del 25	SSA	UK	1.0231
Add25	GA	OK	4.6001	Del 25	GA	OK	5.9211
Add25	GA	UK	4.442	Del 25	GA	UK	5.7212
Add30	SSA	OK	1.0765	Del 30	SSA	OK	1.3593
Add30	SSA	UK	0.8923	Del 30	SSA	UK	1.2322
Add30	GA	OK	4.4211	Del 30	GA	OK	5.9811
Add30	GA	UK	5.8321	Del 30	GA	UK	5.8321

Table 1 Pakistan dataset: adding and deleting measurements to the initial design with OK and UK as interpolation Procedure- SA: Simulated Annealing; GA: Genetic Algorithms.

6 CONCLUSION

The preceding article demonstrates the optimal monitoring network to the rainfall distribution of Pakistan that may be used quite efficiently to other monitoring network as well. Environmental variables are key variables in management of the water resources to every field. These variables are measured concerning the space and location and these should be considered spatially dependent. Among all environmental parameters/variables, precipitation usually consider as one of the vital variables which effects the climate. Identifying the optimal location for the precipitation is a key tool to provide the precipitation's spatial and temporal behavior. Prediction on optimal location can provide the accurate prediction of the precipitation which enable us to make the policies and efforts to manage and reduce the damage because of the drastic events like, flood, droughts etc.

OK and UK are the interpolation techniques to predict the spatial variables (rainfall) on the unobserved locations based on observed locations. We observed that MSPE of the UK has been remained comparatively smaller as compared to the OK by using the cross-validation technique.

Among many of the methods, stochastic search algorithms provide the possibility to identify the optimal locations where we can predict the interest variable more accurately. We utilized and modified the SSA and GA to find the optimal monitoring network to the rainfall monitoring network of Pakistan. From the results, we revealed that these both methods are quite efficient to design an optimal monitoring network of the environmental variables like rainfall in terms of time and less complexity. It is also evidently noted that, the SSA with UK remained efficient for any new optimal monitoring network of rainfall in our case. Furthermore, we notice that when we reduce or expand the optimal network for smaller locations like 1-10 in our case, GA remained efficient in terms of time and converges quickly (maybe in next research someone can consider the time as interest variables during the comparison of any stochastic search algorithm). However, the SSA is possibly the best choice from our finding to select an optimal monitoring network for the rainfall distribution and this result is aligned with the finding of (Brus and Heuvelink, 2007).

DATA ACCESSIBILITY STATEMENTS

Data used in this research is taken from Pakistan Meteorological Department, Islamabad available online at: <https://www.pmd.gov.pk>. Research codes will be provided on personal request.

ABBREVIATIONS

MSPE: Mean Square Prediction Error

GA: Genetic Algorithm

UK: Universal Kriging

OK: Ordinary Kriging

AKV: Average Kriging Variance

SSA: Spatial Simulated Annealing

ETHICS AND CONSENT

This article does not contain any studies with human participants performed by the author.

ACKNOWLEDGEMENT

The authors are thankful to the Editor and anonymous referees for their valuable and constructive comments/suggestions, which certainly improved the presentation and quality of the paper.

FUNDING INFORMATION

The authors declare no funding for the research article.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Talha Omer  orcid.org/0000-0003-4793-9683

Department of Economics, Finance and Statistics, JIBS, Jönköping University, Sweden; Department of Statistics and Computer Science, University of Veterinary and Animal Sciences, UVAS, Lahore, Pakistan

Mahmood Ul Hassan  orcid.org/0000-0003-2889-0263

Department of Statistics, Stockholm University, Stockholm, Sweden

Ijaz Hussain  orcid.org/0000-0002-1586-1503

Department of Statistics, Quaid-I-Azam University, Islamabad, Pakistan

Maryam Ilyas  orcid.org/0000-0002-1662-8340

College of Statistical and Actuarial Sciences, University of the Punjab, Lahore, Pakistan

Syed Ghulam Mohayud Din Hashmi  orcid.org/0000-0001-7696-9414

Department of Wildlife and Ecology, University of Veterinary and Animal Sciences, Lahore, Pakistan

Yousaf Ali Khan  orcid.org/0000-0001-9508-7740

Department of Mathematics and Statistics, Hazara University, Mansehra, Pakistan

REFERENCES

- Adib, A and Moslemzadeh, M.** 2016. Optimal selection of number of rainfall gauging stations by kriging and genetic algorithm methods. *International Journal of Optimization in Civil Engineering*, 6(4): 581–594.
- Al-Mudhafar, WJ.** 2019. Bayesian kriging for reproducing reservoir heterogeneity in a tidal depositional environment of a sandstone formation. *Journal of Applied Geophysics*, 160: 84–102. DOI: <https://doi.org/10.1016/j.jappgeo.2018.11.007>
- Baume, OP, Gebhardt, A, Gebhardt, C, Heuvelink, GB and Pilz, J.** 2011. Network optimization algorithms and scenarios in the context of automatic mapping. *Computers & Geosciences*, 37(3): 289–294. DOI: <https://doi.org/10.1016/j.cageo.2010.04.014>
- Bélisle, CJ.** 1992. Convergence theorems for a class of simulated annealing algorithms on R d. *Journal of Applied Probability*, 29(4): 885–895. DOI: <https://doi.org/10.2307/3214721>
- Bostan, PA, Heuvelink, GB and Akyurek, SZ.** 2012. Comparison of regression and kriging techniques for mapping the average annual precipitation of Turkey. *International Journal of Applied Earth Observation and Geoinformation*, 19: 115–126.
- Brus, DJ and Heuvelink, GB.** 2007. Optimization of sample patterns for universal kriging of environmental variables. *Geoderma*, 138(1–2): 86–95.
- Carrera-Hernández, JJ and Gaskin, SJ.** 2007. Spatio temporal analysis of daily precipitation and temperature in the Basin of Mexico. *Journal of Hydrology*, 336(3–4): 231–249.
- Deutsch, CV and Cockerham, PW.** 1994. Practical considerations in the application of simulated annealing to stochastic simulation. *Mathematical Geology*, 26(1): 67–82. DOI: <https://doi.org/10.1007/BF02065876>
- Diggle, PJ and Ribeiro, PJ.** 2007. Geostatistical design. *Model-based geostatistics*: 199–212. DOI: https://doi.org/10.1007/978-0-387-48536-2_8
- Ellahi, A, Hussain, I, Hashmi, MZ, Almazah, MMA and Al-Duais, FS.** 2021. Agricultural drought periods analysis by using nonhomogeneous poisson models and regionalization of appropriate model parameters. *Tellus A: Dynamic Meteorology and Oceanography*, 73(1): 1–16. DOI: <https://doi.org/10.1080/16000870.2021.1948241>
- Fabian, V.** 1997. Simulated annealing simulated. *Computers & Mathematics with Applications*, 33(1–2): 81–94.
- Fleischer, MA.** 1996. Cybernetic optimization by simulated annealing: Accelerating convergence by parallel processing and probabilistic feedback control. *Journal of Heuristics*, 1(2): 225–246. DOI: <https://doi.org/10.1007/BF00127079>
- Fleischer, MA.** 1999. Generalized cybernetic optimization: Solving continuous variable problems. In *Meta-Heuristics*, 403–418. Boston, MA: Springer.
- Fuentes, M, Chaudhuri, A and Holland, DM.** 2007. Bayesian entropy for spatial sampling design of environmental data. *Environmental and Ecological Statistics*, 14(3): 323–340. DOI: <https://doi.org/10.1007/s10651-007-0017-0>
- Gallagher, K and Sambridge, M.** 1994. Genetic algorithms: a powerful tool for large-scale nonlinear optimization problems. *Computers & Geosciences*, 20(7–8): 1229–1236.
- Goldberg, D.** 1975. “Adaptation in natural and artificial systems”. Ann Arbor, MI: University of Michigan Press.
- Gringarten, E and Deutsch, CV.** (1999, October). Methodology for variogram interpretation and modeling for improved reservoir characterization. In *Spe annual technical conference and exhibition*. OnePetro. DOI: <https://doi.org/10.2118/56654-MS>
- Guedes, LPC, Ribeiro, PJ, Jr, De Stefano Piedade, SÔNIA and Uribe-Opazo, MA.** 2011. Optimization of spatial sample configurations using hybrid genetic algorithm and simulated annealing. *Chilean Journal of Statistics (ChJS)*, 2(2).
- Hengl, T, Heuvelink, GB and Rossiter, DG.** 2007. About regression-kriging: From equations to case studies. *Computers & geosciences*, 33(10): 1301–1315. DOI: <https://doi.org/10.1016/j.cageo.2007.05.001>
- Hengl, T, Heuvelink, GB and Stein, A.** 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma*, 120(1–2): 75–93.
- Heuvelink, GB, Jiang, Z, De Bruin, S and Twenhöfel, CJ.** 2010. Optimization of mobile radioactivity monitoring networks. *International Journal of Geographical Information Science*, 24(3): 365–382. DOI: <https://doi.org/10.1080/13658810802646687>
- Hussain, I, Faisal, M, Shad, MY, Hussain, T and Ahmed, S.** 2015. Assessment of spatial models for interpolation of elevation in Pakistan. *International Journal of Global Warming*, 7(3): 409–422. DOI: <https://doi.org/10.1504/IJGW.2015.069371>
- Hussain, I, Pilz, J and Spoeck, G.** 2011. Homogeneous climate regions in Pakistan. *International Journal of Global Warming*, 3(1–2): 55–66.
- Hussain, I, Shakeel, M, Faisal, M, et al.** 2014. Distribution of Total Dissolved Solids in Drinking Water by Means of Bayesian Kriging and Gaussian Spatial Predictive Process. *Water Qual Expo Health*, 6: 177–185. DOI: <https://doi.org/10.1007/s12403-014-0123-9>
- Hussain, I, Spöck, G, Pilz, J and Yu, HL.** 2010. Spatio-temporal interpolation of precipitation during monsoon periods in Pakistan. *Advances in water resources*, 33(8): 880–886. DOI: <https://doi.org/10.1016/j.advwatres.2010.04.018>
- Ikechukwu, MN, Ebinne, E, Idorenyin, U and Raphael, NI.** 2017. Accuracy assessment and comparative analysis of IDW, spline and kriging in spatial interpolation of landform (topography): an experimental study. *Journal of Geographic Information System*, 9(03): 354. DOI: <https://doi.org/10.4236/jgis.2017.93022>

- Isaaks, EH and Srivastava, RM.** 1989. An Introduction to Applied Geostatistics, Oxford University Press, 561 pages.
- Journel, AG.** 1990. Geostatistics for reservoir characterization. *SPE-20750-MS*, presented at the SPE Annual Technical Conference and Exhibition, (23–26 September), New Orleans, Louisiana.
- Khan, S, Hussain, I and Rahman, A.** 2021. Identification of homogeneous rainfall regions in New South Wales, Australia. *Tellus A: Dynamic Meteorology and Oceanography*, 73(1): 1–11. DOI: <https://doi.org/10.1080/16000870.2021.1907979>
- Kirsch, TD, Wadhwani, C, Sauer, L, Doocy, S and Catlett, C.** 2010. Impact of the 2010 Pakistan floods on rural and urban populations at six months. *PLoS Curr*, 2012 Aug 22; 4: e4fdb212d2432. DOI: <https://doi.org/10.1371/4fdb212d2432>
- Knotters, M, Brus, DJ and Voshaar, JO.** 1995. A comparison of kriging, co-kriging and kriging combined with regression for spatial interpolation of horizon depth with censored observations. *Geoderma*, 67(3–4): 227–246. DOI: [https://doi.org/10.1016/0016-7061\(95\)00011-C](https://doi.org/10.1016/0016-7061(95)00011-C)
- Matheron, G.** 1963. Principles of geostatistics. *Economic geology*, 58(8): 1246–1266. DOI: <https://doi.org/10.2113/gsecongeo.58.8.1246>
- Matheron, G.** 1989. The internal consistency of models in geostatistics. In *Geostatistics*, 21–38. Dordrecht: Springer.
- Metropolis, N, Rosenbluth, AW, Rosenbluth, MN, Teller, AH and Teller, E.** 1953. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6): 1087–1092.
- Molla, A, Zuo, S, Zhang, W, Qiu, Y, Ren, Y and Han, J.** 2022. Optimal spatial sampling design for monitoring potentially toxic elements pollution on urban green space soil: A spatial simulated annealing and k-means integrated approach. *Science of The Total Environment*, 802: 149728. DOI: <https://doi.org/10.1016/j.scitotenv.2021.149728>
- Mubarak, N, Hussain, I, Faisal, M, Hussain, T, Shad, MY, AbdEl-Salam, NM and Shabbir, J.** 2015. Spatial distribution of sulfate concentration in groundwater of South-Punjab, Pakistan. *Water Quality, Exposure and Health*, 7(4): 503–513. DOI: <https://doi.org/10.1007/s12403-015-0165-7>
- Nasseri, M, Asghari, K and Abedini, MJ.** 2008. Optimized scenario for rainfall forecasting using genetic algorithm coupled with artificial neural network. *Expert systems with applications*, 35(3): 1415–1421. DOI: <https://doi.org/10.1016/j.eswa.2007.08.033>
- Omer, T, Hussein, Z and Qasim, M.** 2019. Optimized monitoring network of Pakistan. In *2nd Asia-Pacific Conference on Applied Mathematics and Statistics*, University of Malaya, Kuala Lumpur, Malaysia, February 21–24, 2019.
- Pardo-Igúzquiza, E.** 1998. Optimal selection of number and location of rainfall gauges for areal rainfall estimation using geostatistics and simulated annealing. *Journal of hydrology*, 210(1–4): 206–220. DOI: [https://doi.org/10.1016/S0022-1694\(98\)00188-7](https://doi.org/10.1016/S0022-1694(98)00188-7)
- Phillips, DL, Lee, EH, Herstrom, AA, Hogsett, WE and Tingey, DT.** 1997. Use of auxiliary data for spatial interpolation of ozone exposure in southeastern forests. *Environmetrics: The official journal of the International Environmetrics Society*, 8(1): 43–61. DOI: [https://doi.org/10.1002/\(SICI\)1099-095X\(199701\)8:1<43::AID-ENV237>3.0.CO;2-G](https://doi.org/10.1002/(SICI)1099-095X(199701)8:1<43::AID-ENV237>3.0.CO;2-G)
- Spöck, G and Hussain, I.** 2012. Spatial sampling design based on convex design ideas and using external drift variables for a rainfall monitoring network in Pakistan. *Statistical Methodology*, 9(1–2): 195–210. DOI: <https://doi.org/10.1016/j.stamet.2011.01.004>
- Spöck, G and Pilz, J.** 2010. Spatial sampling design and covariance-robust minimax prediction based on convex design ideas. *Stochastic Environmental Research and Risk Assessment*, 24(3): 463–482. DOI: <https://doi.org/10.1007/s00477-009-0334-y>
- Santacruz, A, Rubiano, Y and Melo, C.** 2014. “Evolutionary optimization of spatial sampling networks designed for the monitoring of soil organic carbon”. In *Soil Carbon*, 77–84. Springer.
- Soroush, F and Abedini, MJ.** 2019. Optimal selection of number and location of pressure sensors in water distribution systems using geostatistical tools coupled with genetic algorithm. *Journal of Hydroinformatics*, 21(6): 1030–1047. DOI: <https://doi.org/10.2166/hydro.2019.023>
- Wadoux, AMC, Brus, DJ, Rico-Ramirez, MA and Heuvelink, GB.** 2017. Sampling design optimisation for rainfall prediction using a non-stationary geostatistical model. *Advances in Water Resources*, 107: 126–138.
- Wang, J, He, T, Lv, C, Chen, Y and Jian, W.** 2010. Mapping soil organic matter based on land degradation spectral response units using Hyperion images. *International Journal of Applied Earth Observation and Geoinformation*, 12: S171–S180.
- Wang, J, Liu, J, Zhuan, D, Li, L and Ge, Y.** 2002. Spatial sampling design for monitoring the area of cultivated land. *International Journal of Remote Sensing*, 23(2): 263–284. DOI: <https://doi.org/10.1080/01431160010025998>
- Zahid, E, Hussain, I, Spöck, G, Faisal, M, Shabbir, J, AbdEl-Salam, NM and Hussain, T.** 2016. Spatial Prediction and Optimized Sampling Design for Sodium Concentration in Groundwater. *PLoS one*, 11(9): e0161810.
- Zhu, Z, and Stein, ML.** 2006. Spatial sampling design for prediction with estimated parameters. *J. Agric. Biol. Environ. Statist.*, 11: 24–44.

TO CITE THIS ARTICLE:

Omer, T, Ul Hassan, M, Hussain, I, Ilyas, M, Din Hashmi, SGM and Khan, YA. 2022. Optimization of Monitoring Network to the Rainfall Distribution by Using Stochastic Search Algorithms: Lesson from Pakistan. *Tellus A: Dynamic Meteorology and Oceanography*, 74(2022), 333–345. DOI: <https://doi.org/10.16993/tellusa.247>

Submitted: 17 July 2022 **Accepted:** 26 July 2022 **Published:** 04 August 2022

COPYRIGHT:

© 2022 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Tellus A: Dynamic Meteorology and Oceanography is a peer-reviewed open access journal published by Stockholm University Press.

