Evaluation and design of a rain gauge network using a statistical optimization method in a severe hydro-geological hazard prone area

Cite as: AIP Conference Proceedings **1836**, 020055 (2017); https://doi.org/10.1063/1.4981995
Published Online: 05 June 2017

Grazia Fattoruso, Antonia Longobardi, Alfredo Pizzuti, Mario Molinara, Claudio Marocco, Saverio De Vito, Francesco Tortorella, and Girolamo Di Francia

Evaluation and design of a rain gauge network using a statistical optimization method in a severe hydro-geological hazard prone area

Grazia Fattoruso[1,a)], Antonia Longobardi[2], Alfredo Pizzuti[2], Mario Molinara[3], Claudio Marocco[3], Saverio De Vito[1,] Francesco Tortorella[3], Girolamo Di Francia[1]

[1] DTE/FSN/DIN *Lab., ENEA Research Center Portici, P.le E. Fermi, 1 80055 Portici Naples, Italy*
[2]*Civil Engineering Department, University of Salerno, 84084 Fisciano, Italy*
[3]*Electromagnetism, Information Engineering and Mathematics Dept., University of Cassino and Southern Lazio, Via G. di Biasio, 43 Cassino, Italy*

[a)]grazia.fattoruso@enea.it

Abstract. Rainfall data collection gathered in continuous by a distributed rain gauge network is instrumental to more effective hydro-geological risk forecasting and management services though the input estimated rainfall fields suffer from prediction uncertainty. Optimal rain gauge networks can generate accurate estimated rainfall fields. In this research work, a methodology has been investigated for evaluating an optimal rain gauges network aimed at robust hydrogeological hazard investigations. The rain gauges of the Sarno River basin (Southern Italy) has been evaluated by optimizing a two-objective function that maximizes the estimated accuracy and minimizes the total metering cost through the variance reduction algorithm along with the climatological variogram (time-invariant). This problem has been solved by using an enumerative search algorithm, evaluating the exact Pareto-front by an efficient computational time.

INTRODUCTION

Rainfall data collection gathered in continuous by a distributed rain gauge network into catchments and into the surrounding areas is instrumental to more effective forecasting and management services to be provided to communities at hydro-geological risk. The rain gauges stations are generally linked to control centers by telemetry for allowing operators to monitor the situation, giving warnings against indicator or trigger rainfall levels, as well as providing inputs into forecasting models.

Rainfall data involved in the numerical models as well as in the alert procedures is commonly characterized as a two-dimensional random field taking into account the time and space variability of the rainfall process in a statistical way. In this framework, the locations and rates along with time interval of sampling are key parameters for accurately estimating the rainfall field. Evaluation of a rain gauges network is essentially based on: (1) number of rain gages and their location (space evaluation) and (2) time interval for measurement (sampling interval) for each rain gauge and duration of measurement (time evaluation). The lack of appropriate information on the rainfall process to be modelled can generate unbefitting estimated rainfall fields, impacting on the ensuing decisions.

In literature, the optimal rain gauges placement problem, has often been formulated as a multi-objective optimization problem. The main objectives to be achieved were to maximize the estimation accuracy of the rainfall process (i.e. information gain adding measurements) and minimize the total metering cost. For quantifying and defining measures of effectiveness in order to assess any given sensor layout, different approaches have been proposed [7]. Among others, the variance reduction analysis is recognized an effective algorithm for the evaluation and design of optimal data collection schemes in random fields.

Applied Mathematics and Computer Science
AIP Conf. Proc. 1836, 020055-1–020055-6; doi: 10.1063/1.4981995
Published by AIP Publishing. 978-0-7354-1506-5/$30.00

This method uses the kriging variance [8-9] as indicator of the level of the estimation uncertainty in the field variable as well as measure for optimal sampling. The area with the highest level of estimation uncertainty is targeted for further monitoring. In order to minimize the estimation uncertainty, a measurement point is added at the location selected on the basis of provided maximum information gain respect with the estimation of the whole field. So, two optimality criteria are utilized for the ranking of potential sampling locations. The first one reflects the amount of information gain (i.e., the variance reduction) due to a new measurement. The second function is proportional to the expected economic gains (i.e., the loss reduction) due to further sampling.

In this research work, a methodology has been investigated for evaluating an optimal rain gauges network aimed at modelling accurate rainfall fields for supporting robust forecasting and management services within severe hydro-geological hazard prone areas such as the case study of the Sarno River basin (Southern Italy).

The optimal rain gauges locations problem within a river basin and into surrounding area has here been formulated as a two-objective optimization problem intended to assess the Pareto front for trade-offs between rain gauges cost and the resulting estimation accuracy of the rainfall filed, quantifying the estimation uncertainty by the variance reduction analysis along with the climatological variogram (time-invariant). The optimization problem has been solved by using an enumerative search algorithm, evaluating the exact Pareto-front by an efficient computational time.

THE METHODOLOGY

The optimal rain gauge network evaluation is here formulated as a two-objective optimization problem under estimation uncertainty. In particular, the objective is to evaluate a distributed rain gauges network aimed at achieving effective estimated rainfall fields within hydro-geological hazard prone areas.

The Problem formulation

The two objectives of the optimization problem (evaluation of an optimal rain gauges network) here investigated are the maximization of the estimation accuracy of the rainfall field variable and the minimization of the sampling cost. Following the traditional literature, the rainfall over a basin is modeled as a two-dimensional random field $P(t,x)$ on $\Re^2$, taking into account the time and space variability of the rainfall process in a statistical way.

Formally, $p(t_j, x_i),\ i = 1, \dots, N$ is the j-th realization, corresponding j-th rainfall event, of the field $P(t, x)$ known in the N discrete locations x_i, where N is the number of rainfall gauges. Considering a catchment area $A \subset \Re^2$ with rainfall measurement stations numbered 1 to N, the rainfall mean of the j-th event over the area A will be $P(t_j) = \frac{1}{A}\int p(t_j, x)dx$.

For our scope, we assume the isotropic and space-stationary hypothesis and consider the optimal (linear, unbiased, minimum variance) estimator of $P(t_j)$, computed for each j from a set of N rainfall observations by $\hat{P}(t_j) = \sum_{i=1}^{N} \lambda_i(t_j) * p(t_j, x_i)$, where the coefficients are the solution of the kriging system [9]:

$$\begin{cases} \displaystyle\sum_{k=1}^{N} \lambda_k(t_j) * \gamma(t_j, h_{ij}) + \mu(t_j) = \frac{1}{A}\int \gamma(t_j, x_i, u)du \\ i = 1, \dots, N \\ \displaystyle\sum_{k=1}^{N} \lambda_k(t_j) \end{cases} \tag{1}$$

where $\mu(t_j)$ is a Lagrange parameter. Discretizing the area A by a square grid of M nodes (numbered N+1 to M) the following numerical assumption is used $\frac{1}{A}\int \gamma(t_j, x_i, u)du \approx \frac{1}{M}\sum_{k=1}^{M} \gamma(t_j, x_i, x_j)\ ,\ x_j \in A$.

With the numerical approximation, the estimation variance is hence given by:

$$\sigma_E^2(t_j) = \mu(t_j) + \sum_{i=1}^{N} \lambda_i(t_j) * \bar{\gamma}(t_j, h_{iA}) - \frac{1}{M^2}\sum_{i=1}^{M}\sum_{k=1}^{M} \gamma(t_j, x_i, x_k)\, x_i, x_k \tag{2}$$

To quantify the estimation accuracy of $P(t_j)$, by $\hat{P}(t_j)$, kriging variance is considered according to the variance reduction method [8]. Consequently, the first optimization objective becomes to minimize the prediction uncertainty given by $\sigma_E^2(t_j)$. Note that $\sigma_E^2(t_j)$ depends on the variogram model, the number N of rain gauges and their spatial locations, and the particular event tj, parameter of the variogram. This means that the variogram should be computed for each realization of the random field in time. However, this drawback can be overcome by calculating the relative variance value between different alternatives for different configurations of sampling data rather that the absolute variance value, and assuming the climatological variogram [1;4;5] as variogram model.

The climatological variogram model is given by $\gamma(\mathrm{t_j}, \mathrm{h}) = \alpha(t_j) * g(h, \beta)$, where $\alpha(t_j)$ is the scaling parameter (space invariant but time dependent), $\beta$is the shape parameter (time invariant but space-dependent), $g(h, \beta)$ can be any variogram model (spherical, exponential, Gaussian, etc.) or the power model used in Bastin et al. (1984), such as $\gamma(t_j, h) = \alpha(t_j) * h^{\beta}$ with $g(h, \beta) = h^{\beta}$and $\beta \in]0{,}2[$.

Thus, the scaled estimation variance becomes

$$\sigma_{\mathrm{E}}^2(\mathrm{t_j}) = \alpha(t_j) v_E^2 \quad and \quad v_E^2 = \mu + \sum_{i=1}^{N} \lambda_i * \bar{g}(h_{iA}) - \bar{g}(h_{AA}) \qquad (3)$$

Since $\sigma_{\mathrm{E}}^2(\mathrm{t_j})$ depends on the scaled climatological variogram (time-invariant), the number N of the rain gauge stations and their locations, the optimal rain gauges placement schemas returned by the described approach can be considered time invariant i.e. effective for each rainfall event.

Thus, given the climatological variogram, the estimation variance $\sigma_{\mathrm{E}}^2(\mathrm{t_j})$ only depends on the number N of the rain gauge stations and their locations. Consequently, it is calculated for each possible placement scheme selecting as optimal, the placement scheme associated to minimum estimation variance value.

So, the first objective intending to maximize prediction accuracy of the rainfall filed variable is obtained with the estimation of the prediction uncertainties. It is hence calculated as:

$$min\, F_1 = \delta(N, x_i) = v_E^2 \qquad (4)$$

The second objective addresses the problem of total metering costs by using a surrogate measure due to the difficulty for defining the capital and operation sensors costs. Assuming for each rain gauge a unitary change in cost, the second objective value is hence calculated as:

$$min\, F_2 = \mathrm{N} \;\; with \;\; N_{min} \leq N \leq N_{max} \qquad (5)$$

where N is the number of rain gauge stations evaluated optimal, N_{min} is the minimum required number of rain gauge stations and N_{max} is the maximum allowed number of rain gauge stations.

The solution of the defined optimization problem is essentially to identify the vectors (N, ν_N^*, Φ_N^*), where for a given N, Φ_N^* is the optimal rain gauge placement scheme with the minimum estimation uncertainty valuated by kriging variance ν_N^*.

Enumerative search solution

The evaluation of an optimal rain gauges network, as formulated by the objective functions (4) and (5), is a complex two-objective optimization problem. The two objective functions are conflicting, so that there exists a number of Pareto optimal solutions to be found. Indeed, there does not exist a single solution that simultaneously optimizes each two objectives, being the accuracy monotonically increasing on the number of sampling data. A solution is called Pareto-optimal if none of the objective functions can be improved in value without degrading some of the other objective values.

The optimal solutions (i.e. sensor placement schemes) to be here found are given by the trade-off between the two conflicting objectives. The Pareto front to be calculated consists in coupled values (N, υ) i.e. the best rainfall estimation accuracy (i.e. minimum estimation variance υ) achieved by a given number N of rain gauges. These pairs form the set of optimal solutions, and a choice can be done on the basis of additional considerations which may be related to the available budget for the maintenance of existing stations as well as the level of estimation accuracy required for the rainfall field, and so on.

The Pareto front calculation can be computationally complex requiring a high computational time, since the evaluation of all possible solutions and the selection of the Pareto-optimal ones may lead to an exponential search space (or infinite in the case of continuous objective functions). For overcoming this drawback, in the literature, several strategies have been proposed, essentially based on evolutionary algorithms, that obtain an adequate approximation of the Pareto front at an acceptable computational cost. Being discrete the formulated problem and limited the range of both the objectives (sensors number and estimation accuracy value), an alternative approach that the genetic algorithms has here been experimented based on the enumerative search method [3]. This approach consists in a conceptually simple search strategy to solve multi-objective optimization problems based on the evaluation of each possible solution in a finite search space properly discretized. Although this approach can become unfeasible for larger search space, it allows obtaining interesting solutions at the multi-objective problem.

THE CASE STUDY

The optimal rain gauge network evaluation problem investigated in this research work, has been applied to the Sarno River basin, located in Campania (Southern Italy).

The Sarno river basin has an extension of about 400 km^2 and is one of most important economic areas in the Campania region, with numerous industrial and agricultural settlements. Urban and peri-urban areas cover a significant portion of total area, especially in the in the valley plains. Pyroclastic deposits generally cover the calcareous-dolomitic rocks and this makes the area particularly prone to debris flow initiations which, along with floodplain flooding, are one of the major sources of hydro-geological damaging natural events for the studied catchment [6].

Problem description

The Sarno river basin reference rain gauge network, including catchment surrounding areas, consist of 30 stations (Figure 1). Earlier managed by the Servizio Idrografico e Mareografico Nazionale, since 2000 it is managed by the Civil Protection Service of the Campania Region. A set of 30 rain gauge stations, located within the catchment area, has been analyzed in the current study. Among these, a subset, identified by progressive indices in the interval 1-11, is referred as the historical network, as for them the monitoring period starts in 1920. For the remaining subset, identified by progressive indices in the interval 12-30, the monitoring period starts in 2002, as they were only recently installed to integrate the existing rain gauge network, in order to improve the spatial representation of precipitation fields.

For each station, the average maximum annual 24 hours precipitation data have been analyzed as they are relevant for hydrogeological hazard investigation and as they represent the least impacted time series by missing data for the considered catchment [2].

The aim of this study case has been to evaluate the existing rain gauges over the Sarno river basin according to an optimally criteria aimed at providing effective estimated rainfall fields that capturing properly the space-time variability of the rainfall events allows to provide more robust forecasts and analysis in severe hydro-geological hazard prone areas.

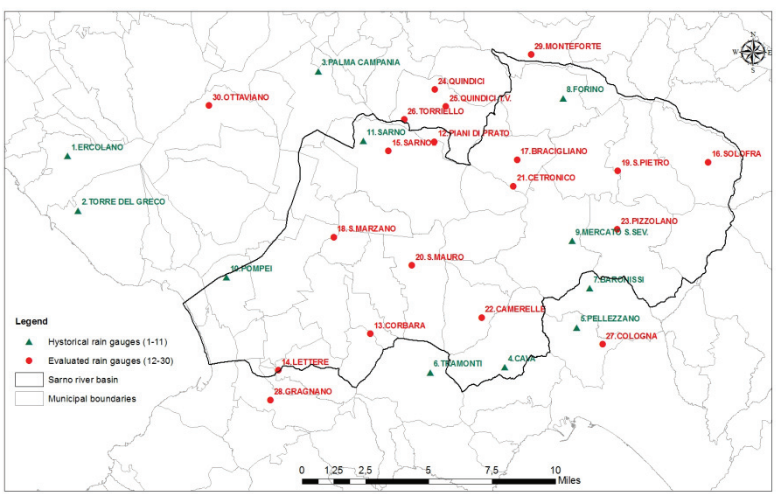

FIGURE 1: Map of the Sarno river basin reference rain gauges network with the stations indexed in the interval 1-30

Method application

The optimal rain gauge network on the Sarno river basin has been evaluated by maximizing the prediction accuracy of the estimated rainfall field and minimizing the number of the rain gauges station involved.

The optimization problem has been formulated by means of the objective functions (4) and (5). The concept of scalar optimality has here been replaced by the Pareto optimal one, so that the optimization process computes the

Pareto front given by the set $\{(\mathrm{N},\ \nu_N^*), \mathrm{con}\ \mathrm{N} = 1, \dots, 19\}$, where the Pareto-optimal solution $(\mathrm{N},\ \nu_N^*)$ is selected by the set of the prospective N-rain gauges placement schemes that returns the minimum estimation variance value ν^*.

According to the problem formulation (see the above section), for evaluating the estimation variance ν, the first step is essentially the evaluation of the Kriging system, given by:

$$\begin{cases} \sum_{k=1}^{T} \lambda_k * h_{ij}^{\beta} + \xi = \frac{1}{M} \sum_{S=1}^{M} h_{is}^{\beta} \\ i = 1, \dots, T \\ \sum_{k=1}^{N} \lambda_k(t_j) \end{cases}$$

where λ_k are the coefficients to be calculated, ξ the Lagrange multiplier, M is the number of samples used for estimating the Bastin's climatological variogram. The T value is the number of sampling points of the filed variable. In particular, it is here given by T= 11+N including the subset of eleven historical rain gauges (i.e. data points), and the subset of N prospective sampling points (i.e optimal N-rain gauges scheme to be added), selected from the subset of 19 rain gauges to be evaluated.

As area of interest of the filed variable (i.e. the catchment area A on the which rainfall filed has to be estimated) has been created a rectangular polygon by the vertices {A(447154.218,4504256.242), B(487854.215,4504256.242), C(487854.215,4526211.146), D(447154.218,4526211.146)}, according to the UTM33/ED50 coordinate system. The rainfall field has been represented by a raster (regular grid) characterized by a spatial resolution of 2 km.

For evaluating the Bastin's climatological variogram, the average maximum annual 24 hours precipitation data on 10 years have been analyzed, gathered by the historical rain gauges. So, the time-invariant parameter β of the Bastin's model variogram has been set to the value 0.728.

Once evaluating the kriging system , for each possible N-rain gauge scheme, the estimation variance has been calculated by $\nu = \xi + \sum_{i=1}^{T} [\lambda_i \cdot \bar{g}(h_{iA}, \beta)] - \bar{g}(h_{AA}, \beta)$ where $\bar{g}(h_{iA}, \beta) = \frac{1}{M} \sum_{s=1}^{M} {h_{is}}^{\beta}$

Thus, for evaluating the Pareto front, the following algorithm has been performed:

1. for each value N=1,...,19 , calculate the $\binom{\mathrm{N}}{19} = \frac{19!}{(19-\mathrm{N})! \cdot \mathrm{N}!}$. possible rain gauges placement schemes, each one named Φ_{Ni},
2. For each N, evaluate the Pareto optimal vector (N, ν_N^*) i.e. for each Φ_{Ni}
   a. evaluate Kriging system where T=11+N data points
   b. evaluate the associated estimation variance ν_{Ni}
   c. evaluate ν_N^* =min ν_{Ni}

The algorithm output is a table of elements $(\mathrm{N}, \nu_N^*, \Phi_N^*)$, where each Φ_N^* rain gauge locations scheme represents the optimal solution with ν_N^* associated estimation variance value. Note that the optimal locations schemes Φ_N^* are shown as sequence of 1 and 0 where 1 means point to be sampled; 0 means point not sampled.

By the Pareto front graph that plots the calculated ν_N^* values vs the N=1,...,19 values, it is possible to visualize the relationship among estimation variance values and the number of evaluated optimal rain gauges.

RESULTS AND DISCUSSION

The optimal Pareto front obtained by the use of the proposed optimal sampling evaluation algorithm is illustrated in Figure 2 for the rain gauge network under investigation. In this graph, on the x-axis the number N of rain gauges to be evaluated is plotted, and on the y-axis the corresponding normalized ν_N^* values, i.e. the minimum estimated variance values corresponding to the optimal locations schemes obtained for N stations.

Clearly, the normalized variance ν_N^* value decreases as the number N of stations included in the optimal rain gauge deployment scheme Φ_N^* increase. The rate of variance reduction is however variable. As an example, a considerable reduction, of about 50%, is approached for N =3, whereas a reduction of about 90% is approached for N = 10. Larger N values further slowly reduce the normalized estimation variance, but their contribution can be assumed negligible in an optimal station deployment scheme.

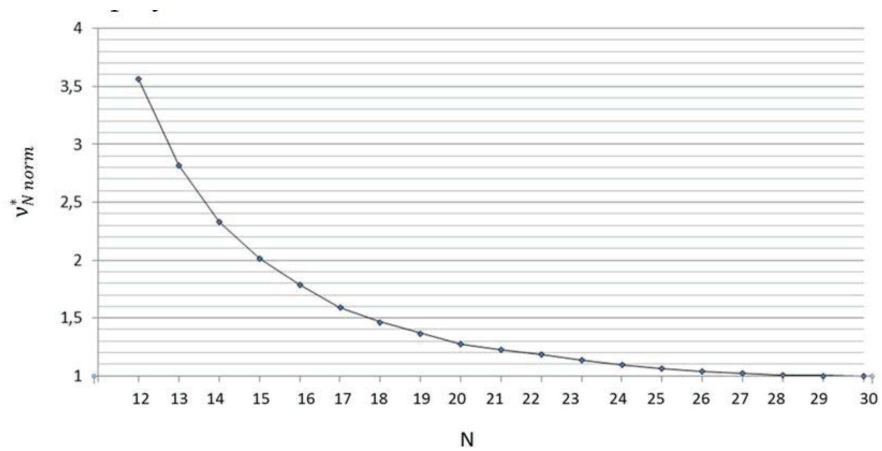

FIGURE 2: Pareto optimal front obtained by the proposed optimal sampling evaluation algorithm

In the Table 1, for each Pareto optimal solution, the related optimal rain gauges locations scheme Φ^*_N is specified as sequence of 0,1 where 0 means point of the rainfall field to be not sampled and 1 to be sampled. Analyzing the sequences, we can notice that the optimal solutions, increasing N, are conservative until to 11-rain gauges placement scheme. In this scheme, the station indexed 25 is replaced by the near station 24 along with the station indexed 12 in order to cover a wider area.

TABLE 1. Pareto optimal solutions

N	$\nu^*_{N\,norm}$	Φ^*_N - id rain gauges																		
		12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	3.5654	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	2.8140	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
3	2.3268	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
4	2.0128	0	0	1	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0
5	1.7893	0	0	1	0	1	0	1	0	1	0	0	0	0	1	0	0	0	0	0
6	1.5915	0	0	1	0	1	0	1	1	1	0	0	0	0	1	0	0	0	0	0
7	1.4648	0	0	1	0	1	0	1	1	1	1	0	0	0	1	0	0	0	0	0
8	1.3683	0	0	1	0	1	0	1	1	1	1	0	0	0	1	0	0	0	1	0
9	1.2766	0	1	1	0	1	0	1	1	1	1	0	0	0	1	0	0	0	1	0
10	1.2275	0	1	1	0	1	0	1	1	1	1	0	1	0	1	0	0	0	1	0
11	1.1833	1	1	1	0	1	0	1	1	1	1	0	1	1	0	0	0	0	1	0
12	1.1391	1	1	1	0	1	0	1	1	1	1	0	1	1	0	0	0	1	1	0
13	1.0973	1	1	1	0	1	0	1	1	1	1	1	1	1	0	0	0	1	1	0
14	1.0631	1	1	1	0	1	0	1	1	1	1	1	1	1	0	0	1	1	1	0
15	1.0399	1	1	1	0	1	0	1	1	1	1	1	1	1	0	0	1	1	1	1
16	1.0230	1	1	1	0	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1
17	1.0082	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1
18	1.0027	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1
19	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

The numerical solution provided by the optimization algorithm is coherent with the requirement for a monitoring network focused on the need to better describe the spatial distribution of rainfall fields for catchment areas frequently hit by natural hydro-geological disasters. Rain gauge station locations that significantly reduce the normalized kriging variance (N up to 10) are indeed located along the main natural hydraulic network and in relieves areas where landslide events, causing losses of lives, occurred in the past years.

Rain gauge stations corresponding to the optimal schemes with N larger than 10, correspond either to locations outside the catchment area or to redundant point measurements located in landslide prone areas.

REFERENCES

1. G Bastin, C. D. Lorent, and M. Gevers, Water Resour. Res., 20, 463–470, (1984)
2. F.Califano, M. Mobilia, A.Longobardi, Procedia Engineering, **119** (1), 1129-1138, (2015)
3. G. Fattoruso et al, Procedia Engineering, **119**, 927-936, (2015)
4. I. F. D.Grimes, Pardo-Iguzquiza Eulogio Geographical Analysis **42**, 136–160, (2010)
5. T. Lebel, G. Bastin, C. Obled, J.D. Creutin,. Water Resour. Res., **23** (11), 2123–2134, (1987)
6. A Longobardi, N Diodato, M Mobilia,. Water (Switzerland, 2016), 8 (9), art. no. 398.
7. A. K Mishra and P. Coulibaly, Rev. Geophys, 47, RG2001, 1-24 (2009).
8. S. Rouhani, Water Resources Research, **6**, 837-846, (1985)
9. R.M Srisvastava, E.H Isaaks,. Practical Geostatistics, Oxford University Press, Oxford.1992