Geostatistically based optimization of a rainfall monitoring network extension: case of the climatically heterogeneous Tunisia

Haifa Feki, Mohamed Slimani and Christophe Cudennec

ABSTRACT

Rainfall data are an essential input for many simulation models. In fact, these latter have a decisive role in the development and application of rational water policies. Since the accuracy of the simulation depends strongly on the available data, the task of optimizing the monitoring network is of great importance. In this paper, an application is presented aiming at the evaluation of a precipitation monitoring network by predicting monthly, seasonal, and interannual average rainfall. The method given here is based on the theory of the regionalized variables using the well-known geostatistical variance reduction method. The procedure, which involves different analysis methods of the available data, such as estimation of the interpolation uncertainty and data cross validation, is applied to a case study data set in Tunisia in order to demonstrate the potential for improvement of the observation network quality. Root mean square error values are the criteria for evaluating rainfall estimation, and network performance is discussed based on kriging variance reduction. Based on this study, it was concluded that some sites should be dropped to eliminate redundancy and some others need to be added to the existing network, essentially in the center and the south, to have a more informative network.

Key words | geostatistics, kriging variance, optimization, rainfall network, spatial variability, Tunisia

INTRODUCTION

Rainfall is highly dynamic and constantly changing in form and intensity across a given region, especially in semi-arid regions (Cudennec et al. 2007). An important stage in many meteorological applications, climatic analysis and engineering design projects is the point or areal estimation of climatic variables over a catchment area from data taken at several weather stations. Traditionally, rainfall rates are derived from rain-gauge measurements. To estimate precipitation properly, an optimal spatial distribution of rain-gauges is required (Xu et al. 2013; Al-Mukhtar et al. 2014; Asghari & Nasseri 2015; Wu et al. in press). Thus, special attention is required not only during the actual measurements but also in the design of the rain-gauge monitoring network.

Rain-gauge monitoring networks are often installed to provide measurements that characterize the temporal and spatial variations in rainfall. However, even though modern rain-gauges are capable of providing the rainfall rate in real time at a very fine resolution in time, the spatial variation of rainfall is still difficult to characterize without a well-designed rain-gauge network (Xu et al. 2015). Pardo-Iguzquiza (1998) has indicated that the problem of rain-gauge networks’ optimization could be considered rather out of date, as current weather radar analysis provides an estimation of rainfall rates at an excellent spatial and temporal resolution. However, complete coverage by weather radar is still limited in some countries and, anyhow, remains dependent on assimilation with ground measurements, all the more in mountainous semi-arid contexts. On the other hand, although recent advances in satellite remote sensing seem to have the potential for providing full spatial coverage...
of pixel rainfall estimates (Chang et al. 1998; Wu et al. 2015), satellite images alone still cannot provide accurate rainfall estimates at the spatial resolution required to match rain-gauge measurements.

Otherwise, the selection of rain-gauge locations is affected by many factors, such as accessibility, ease of maintenance, and topographical aspects, and the minimum density required for a rain-gauge network is dependent on the temporal resolution of the desired rainfall measurements. The optimization of a rain-gauge network for large time scale (annual, monthly, daily, etc.) measurements does not imply that it is optimal for less important ones. Generally, it is not due to the position of the rain-gauges, but due to their number (Pardo-Igúzquiza 1998). A much denser network is often necessary to characterize rainfall patterns with great variability and intermittency, even if optimization is now sometimes used for cost-driven reduction of the networks.

Determining a precise methodology for both expanding and evaluating the existing rain-gauge network is essential to understand the capacity of the network and the quality of the data it provides. Consequently, when selecting rain-gauges, one must take into account both their number and location to ensure the greatest accuracy in estimating rainfall. However, increasing the number of sampling points may produce a redundancy in the data which in turn reduces cost efficiency. Cost considerations are especially critical in developing countries that face extreme pressures on both their water and financial resources.

Designing accurate rainfall monitoring networks has occupied researchers for many years. Over the last decade, a number of public agencies whose purpose is to monitor environmental variables (Nunes et al. 2004; Pardo-Igúzquiza & Dowd 2004; Baalousha 2010) and groundwater variables (Carrera et al. 1984; Prakash & Singh 2000; Theodossiou & Latinopoulos 2006; Yang et al. 2008) have invested great economic, technical, and human resources in planning and operating improvements on existing monitoring networks. Meteorological monitoring networks, and particularly those devoted to rainfall monitoring, are among those which have received most attention from researchers (Kassim & Kottega 1991; Papanicolaou & Metaxa 1996; Ashraf et al. 1997; Pardo-Igúzquiza 1998; Goovaerts 2000). Recent scientific literature has provided various approaches, characterized by different levels of complexity, capable of supporting optimal network design (Barca et al. 2008; Cheng et al. 2008; Chebbi et al. 2011, 2013; Shagaghian & Abedini 2015; Shafiei et al. 2014; Adhikary et al. 2015; Xu et al. 2015; Wu et al. in press). Among others, one type of approach is a kriging-based geostatistical approach that finds wide applications in rain-gauge network design. Optimal network configuration can be achieved by the variance reduction method (Adhikary et al. 2015). In many studies, this technique was employed alone for the rain-gauge network design (Kassim & Kottega 1991; Papanicolaou & Metaxa 1996; Ashraf et al. 1997; Prakash & Singh 2000; Pardo-Igúzquiza & Dowd 2004; Theodossiou & Latinopoulos 2006; Cheng et al. 2008; Chebbi et al. 2013; Shafiei et al. 2014; Adhikary et al. 2015). However, some studies applied the kriging technique in combination with other techniques, such as entropy (Chen et al. 2008; Yeh et al. 2011), multivariate factor analysis (Shagaghian & Abedini 2015), and simulated annealing (Pardo-Igúzquiza 1998; Barca et al. 2008; Chebbi et al. 2011) for the network design. For more details on geostatistical theory, readers can refer to Chiles & Delfiner (1999), Deutsch & Journel (1992), Feki et al. (2012) and Hudson & Wackernagel (1994).

Several important studies have been conducted in the field of spatial analysis of network design. Kassim & Kottega (1991) compared simple and disjunctive kriging in the estimation of optimum locations of rain-gauges as part of a network for the determination of storm characteristics to be used in forecasting and design. Papanicolaou & Metaxa (1996) applied ordinary and universal kriging (UK) to describe the spatial variability of monthly rainfall data, and they used the property of kriging variance for the optimal design of the rain-gauge network. A criterion using ordinary kriging variance is proposed by Cheng et al. (2008) to assess the accuracy of hourly and annual rainfall estimation using the acceptance probability defined as the probability that estimation error falls within a desired range. Chebbi et al. (2011) proposed a method for robust rain-gauge network optimization using intensity–duration–frequency data by minimizing the mean spatial kriging variance. Adhikary et al. (2015) have developed a methodology to design an optimal rain-gauge network through optimal positioning of additional stations as well as optimal relocating of existing redundant stations using...
the kriging-based geostatistical approach for daily rainfall in Australia. Pardo-Igúzquiza (1998) presented a method that defines the optimal network by combining kriging and simulated annealing for daily and annual rainfall. The methodology presented by Barca et al. (2008), which uses geostatistics and simulated annealing, has been applied to the design of an extension to a rainfall monitoring network. A seasonal aggregation of the available rainfall data was considered. Chebbi et al. (2015) combined the variance reduction method with simulated annealing to optimally extend a rain-gauge network in order to interpolate rainfall intensity and an erosivity index. A methodology for the design of a rain-gauge network was developed by Shaghaghian & Abedini (2015) using a combination of geostatistical tools and factor analysis for mean annual rainfall. Shafiei et al. (2017) evaluated the performance and augmentation of a rain-gauge network for annual time scale using a probabilistic approach combined with a geographic information system. Recently, in their study, Wu et al. (in press) intended to treat the radar grid-based data as the reference data set to propose a framework for evaluating the rain-gauge network. The evaluation framework is developed based on quantifying the fitness performance of the semi-variogram of rainfall characteristics.

In summary, there are numerous inter-related factors affecting rain-gauge network design as the objective of designing a network: the process considered, the attribute under consideration, the temporal scale, the spatial scale, the topographic setting, types of precipitation, the nature of the objective function used for and the optimization algorithm used for minimization or maximization purposes (Shaghaghian & Abedini 2015). The main objective of this article is to evaluate an existing rain-gauge network in a semi-arid region using the kriging variance reduction method for three time scales: monthly, seasonal, and interannual. The advantage of the kriging-based approach is that the variance of estimation is independent of the actual measurements; therefore, the network can be pre-designed. We first describe the topographic and rainfall of the study area and the selected data used, before presenting the methodology. Then, the results are given and discussed. The final section gives the conclusions drawn from this research.

MATERIALS AND METHODS

The case study and data

Located to the south of the Mediterranean Sea with an area of 164,150 km², Tunisia displays the best available meteorological observation network and data set of the southern Mediterranean area, in terms of density, history, coherence, and quality at the national level (Slimani et al. 2007; Baccour et al. 2012a, 2012b; Feki et al. 2012), allowing various meteorological applications and climatological assessments. Yet, this network remains far from ideal to capture all the spatial and temporal variables across a strong north–south, orography-driven and seasonally dependent organization. The Tunisian territory has a quite heterogeneous relief (Figure 1). Small
plains where the Wadi Medjerda flows (major permanent river of North Africa) separate mountains in the north. They are progressively replaced by steppes, then by oriental coasts that stretch from the Cap Bon peninsula to the Libyan border. The coasts (1,300 km) are cut by profound gulfs and several islands. Towards the south, the mountainous chains are lower (Sahara). Climatologically, Tunisia is situated in the geographic transition zone between the humid temperate climate and the arid Saharan climate (Feki et al. 2022). Although the global climate is arid, there are three distinct climatic regions: (1) the north region is characterized by a Mediterranean climate with a humid and sub-humid atmosphere; (2) a semi-arid climate is found in the center and the east coast of Tunisia; and (3) the climate becomes Saharan in the south. Precipitation is characterized by an important north–south spatial distribution. In the north, the cumulative annual mean rainfall ranges from 400 mm, 1,000 mm, to 1,500 mm per year over the Kroumirie Mountains. The rainy season extends from September to April with a maximum rainfall of 234.9 mm during the winter months. In the center of Tunisia, mean rainfall ranges from 200 to 400 mm per year and is characterized by its variability from one year to another. South Tunisia is characterized by a Saharan climate where rainfall is erratic.

Monthly rainfall data for 140 rain-gauges in Tunisia (Figure 1) were collected from the General Direction of Water Resources in Tunisia of the Ministry of Agricultural (Direction Générale des Resources en Eau (DGRE)) and the National Institute of Meteorology (Institut National de la Météorologie (INM)). The analysis was performed on annual and seasonal precipitation totals as defined by climatological seasons: spring (March, April, May), summer (June, July, August), autumn (September, October, November), and winter (December, January, February). Then, monthly, seasonal, and interannual average rainfalls were calculated for each rain-gauge station with its own period of observation which varies from 30 to 100 years. Stationarity was confirmed for all rainfall series as they exceed 30 years (Feki 2010). All rainfall amounts were expressed as 1/10 mm and corresponding statistical analysis was carried out in Table 1.

Methodology

Kriging is a spatial interpolation method which predicts distribution of precipitation optimally (Abedini et al. 2013). One
of the most interesting features of kriging, in general, is that the estimation variance can be calculated before actual measurements are available. This feature suggests its application for designing networks of monitoring stations (Carrera et al. 1983). The kriging variance depends on the structure and the geometric configuration of the data points. The further the point estimate is from its neighbors, the more the kriging variance increases. As the kriging weights and the semi-variogram between the estimated point and its neighbors are deducted from the theoretical model of the semi-variogram, the variance of the error depends on the distance between the points and their dispositions in space, regardless of values.

A rain-gauge network encompasses a number of gauges within an area of interest. Two major concerns for the evaluation and design of a network are the density and locations of these rain-gauges. The minimum density of a network depends on its functional objective and characteristics of dominant precipitation types. For water resources management (reservoir operation, water balance, etc.), estimates of long duration rainfall, such as mean monthly, seasonal, and annual rainfall, are needed (Chen et al. 2008; Shaghaghian & Abedini 2013). In evaluating and designing a rain-gauge network, the major concern is how to quantify the performance of an existing network. Intuitively, a measure of the network performance is the accuracy associated with rainfall estimates. The accuracy should reflect not only the absolute error, but also the error variance, since the errors depend heavily on rain field variability, and this variability is far from being consistent from one event to another and from one year to the next. If only unbiased estimators are considered, then the characteristics of estimation accuracy can thus be focused on the variance of estimation error (Cheng et al. 2008). The estimation accuracy of point rainfall varies within the study area and is dependent on the number and geometric layout of rain-gauges.

In this study, evaluation of the rainfall network passes through two major steps: (1) redundancy elimination and (2) network augmentation. To eliminate the redundant rain-gauges in the network, cluster analysis is used. Cluster analysis for hydrologic regionalization involves the grouping of observations or variables into clusters, with each cluster containing observations or variables with highly similar hydrologic characteristics, such as geographical, physical, statistical, or stochastic features (Soltani & Modarres 2006). Ward’s method was used. It is an agglomerative clustering technique, where clusters are grouped together to include increasing numbers of objects in a stepwise fashion. At the lowest level of the clustering hierarchy, each cluster contains a single object. Larger clusters are formed by merging the two clusters that produce the smallest increase in the within-cluster variance. This merger step is repeated until a single cluster containing all objects is created. Ward’s method has been found to produce satisfactory results for meteorological data in previous research (Alhamed et al. 2002). The progress and intermediate results of a cluster analysis are conventionally illustrated using the dendogram or ‘tree’ diagram, a bi-dimensional figure that represents the sequence and the distance at which the observations are clustered. In this case, the y axis indicates decreasing levels of similarity, and ‘branches’ on the dendogram indicate the level of similarity where objects are grouped into clusters.

The augmentation of the number of stations is based on geostatistical methods: UK (Papamichail & Metaxa 1996; Hengl et al. 2004; Brus & Heuvelink 2007) and kriging with external drift (KED) (Goovaerts 2000; Bourennane & King 2003; Feki et al. 2012) algorithms are compared to interpolate data from the monitoring network and to get rainfall maps and the kriging variance maps. According to Hengl et al. (2004), UK is a spatial prediction technique that jointly employs correlation with auxiliary maps and spatial correlation. Optimization of the sample pattern by minimizing the UK variance integrates optimization in feature and geographic space in one step. The method has no ambiguities and does not involve subjective choices (Brus & Heuvelink 2007). Many authors (Deutsch & Journel 1992; Hudson & Wackernagel 1994), however, agree that the term UK should be reserved for the case where the drift is modeled as a function of the coordinates only. The KED is more commonly used when the drift is defined ‘externally’ through some auxiliary variables (Chiles & Delfiner 1999; Feki et al. 2012). The advantage of KED is that the equations are solved at once. Note that, although the KED technique seems to be computationally more straightforward, it needs semi-variogram parameters of the global least square (GLS) regression residuals (which is often ignored), and therefore the GLS regression coefficients.
Many researchers make different assumptions to escape the estimation of residuals’ semi-variograms. For example, Marcotte (2012) assumes that for a small amplitude drift the semi-variogram of the main variable can be used directly at close range. Hudson & Wackernagel (1994) and Bourennane & King (2005) go further by directly assuming that the semi-variogram of the residuals and of the main variable are equal.

Our approach particularly uses the kriging standard deviation to identify points of high variance as potential points for monitoring. The drawback of this approach is that it only considers the spatial variation of the data without considering the accessibility to the region. In other words, if we consider an additional point \( s_i \) to a set of \( n \) points \( s_i(i=1,\ldots,n) \) it is possible to construct the kriging system corresponding to the \( n+1 \) points \( s_1,s_2,\ldots,s_n,s_{n+1} \), and deduce the new variance kriging related to this new experimental configuration. This method is called ‘method of fictitious point’.

Let \( z(s_i) \), which corresponds to rainfall, represent realization of random variable \( Z(s_i) \) at particular points \( s_i \) within a field \( S \). The intrinsic hypothesis assumes that for a random variable \( Z(s_i) \): (i) the expected value of \( Z(s_i) \) does not depend on the position \( s_i \), and (ii) the variance of \( Z(s_i) - Z(s_i + h) \) does not depend on the position \( s_i \) in \( S \) for any separation vector \( h \). The semi-variogram function \( \gamma \) gives a measure of the spatial autocorrelation of a random variable \( Z \) as a function of separation distance \( h \). A semi-variogram is defined as one half of the variance between any two locations separated by \( h \) (Shafiei et al. 2014):

\[
\gamma(h) = (1/2) \text{Var}[Z(s_i + h) - Z(s_i)]
\]

where \( h \) is the distance vector and \( s \) is the location vector.

Equation (1) indicates that the semi-variogram is independent of spatial locations; it only depends on the distance between the two points under consideration in the isotropic case. An experimental semi-variogram is computed from data pairs of observations, for specific distance lags and directions. A mathematical authorized model may be fitted to the experimental semi-variogram, and the coefficients of this model are used for kriging. In semi-variogram models where the semi-variance reaches a finite variance and levels out, the maximum is referred to as the sill. This is termed a bounded model (spherical and exponential models). Models that do not reach a sill are termed unbounded (power models). Unbounded semi-variograms may indicate a large-scale trend (that is, systemic increase or decrease) in the values of the property characterized across the region of study (Kitanidis 1997).

The UK (Brus & Heuvelink 2007) system is given by the following equation:

\[
\sum_{j=1}^{n} \lambda_j \gamma(s_i, s_j) + \sum_{l=0}^{L} \mu_l f_l(s_i) = \gamma(s_i, s_0) \text{ for } i = 1, 2, \ldots n
\]

\[
\sum_{j=1}^{n} \lambda_j f_l(s_j) = f_l(s_0) \text{ for } l = 0, 1 \ldots L
\]

and its estimation variance is:

\[
\sigma_{nk}(s_0) = \sum_{i=1}^{n} \lambda_i \gamma(s_i, s_0) - \sum_{l=0}^{b} \mu_l f_l(s_0)
\]

where \( \mu_l \) are Lagrange parameters, \( f_l (l = 0, 1, 2, \ldots, L) \) are deterministic functions of the geographic coordinates, \( \lambda_j \) are the kriging weights and \( \gamma(h) \) is the semi-variogram of \( Z(s_i) \).

The KED (Bourennane & King 2003; Feki et al. 2012) system is:

\[
\sum_{j=1}^{n} \lambda_j \gamma(s_i - s_j) - \mu_1 - \mu_2 Y(s_i) = \gamma(s_i - s_0) \text{ for } i = 1, \ldots, n
\]

\[
\sum_{j=1}^{n} \lambda_j = 1
\]

\[
\sum_{j=1}^{n} \lambda_j Y(s_j) = Y(s_0)
\]

where the minimal prediction variance:

\[
\sigma_{KED}^2(s_0) = \gamma(0) - \sum_{i=1}^{n} \lambda_i \gamma(s_i - s_0) - \mu_1 - \mu_2 Y(s_0)
\]

where \( \mu_1 \) and \( \mu_2 \) are the Lagrange parameters, \( f_l (l = 0, 1, 2, \ldots, L) \) are deterministic functions of the geographic coordinates, \( \lambda_j \) are the kriging weights, and \( \gamma(h) \) is the semi-variogram of \( Z(s_i) \).
RESULTS AND DISCUSSION

Redundancy elimination

The cluster analysis was performed based on Ward’s method and using the interannual, seasonal, and monthly average rainfall (5 + 12 variables) of the selected rainfall stations. The rain-gauges’ classification is carried out according to different scales. In fact, there are, initially, two big clusters which represent the north and the south of the country. Then, when the dissimilarity decreases, the number of clusters is more and more important. For this study, we have retained clusters corresponding to the third level of dissimilarity. This classification divided the rain-gauges set into 16 clusters with similar climatic and geographic characteristics. The matrix distance between the 140 rainfall stations is calculated. Then, the series of rain-gauges within 10 km distance of each other (Ferguson 1972) are compared and stations with more complete rainfall series are retained for each cluster. Finally, 101 stations were retained (Figure 2).

Characterizing the spatial variability of rainfall

The 101 rain-gauges are used in the semi-variogram calculation using the VARIOWIN Software. Under the

<table>
<thead>
<tr>
<th>Rainfall simple semi-variogram</th>
<th>Model</th>
<th>(c_0)</th>
<th>(a^*)</th>
<th>(b^{**})</th>
</tr>
</thead>
<tbody>
<tr>
<td>September</td>
<td>Power</td>
<td>0</td>
<td>0.07</td>
<td>1</td>
</tr>
<tr>
<td>October</td>
<td>Power</td>
<td>5,478</td>
<td>0.26</td>
<td>1</td>
</tr>
<tr>
<td>November</td>
<td>Power</td>
<td>10,780</td>
<td>0.5</td>
<td>0.98</td>
</tr>
<tr>
<td>December</td>
<td>Power</td>
<td>34,200</td>
<td>0.68</td>
<td>1</td>
</tr>
<tr>
<td>January</td>
<td>Power</td>
<td>17,600</td>
<td>0.8</td>
<td>0.99</td>
</tr>
<tr>
<td>February</td>
<td>Power</td>
<td>18,430</td>
<td>0.48</td>
<td>0.98</td>
</tr>
<tr>
<td>March</td>
<td>Power</td>
<td>9,870</td>
<td>0.3</td>
<td>0.96</td>
</tr>
<tr>
<td>April</td>
<td>Power</td>
<td>6,290</td>
<td>0.2</td>
<td>0.98</td>
</tr>
<tr>
<td>May</td>
<td>Power</td>
<td>1,120</td>
<td>0.07</td>
<td>1</td>
</tr>
<tr>
<td>June</td>
<td>Power</td>
<td>312</td>
<td>0.03</td>
<td>0.98</td>
</tr>
<tr>
<td>July</td>
<td>Spherical</td>
<td>73</td>
<td>730</td>
<td>210,548</td>
</tr>
<tr>
<td>August</td>
<td>Power</td>
<td>594</td>
<td>0.018</td>
<td>0.98</td>
</tr>
<tr>
<td>Autumn</td>
<td>Power</td>
<td>24,586</td>
<td>1.78</td>
<td>1</td>
</tr>
<tr>
<td>Winter</td>
<td>Power</td>
<td>221,000</td>
<td>6.5</td>
<td>0.98</td>
</tr>
<tr>
<td>Spring</td>
<td>Power</td>
<td>47,600</td>
<td>1.14</td>
<td>1</td>
</tr>
<tr>
<td>Summer</td>
<td>Power</td>
<td>690</td>
<td>0.2</td>
<td>0.96</td>
</tr>
<tr>
<td>Interannual</td>
<td>Power</td>
<td>572,000</td>
<td>22.9</td>
<td>1</td>
</tr>
</tbody>
</table>

\(c_0\) (1/10 mm²) nugget effect; \(a^*\) (1/10 mm²) sill (for the spherical model) or slope (for the power model); \(b^{**}\) (1/10 m) range (for the spherical model) or power coefficient (for the power model).
second order assumption, a semi-variogram approaches its sill. However, except for July, the remaining calculated semi-variograms are fitted visually by power type (Table 2). Therefore, the semi-variograms increase without a sill; it means that the random rainfall field is non-stationary due to a spatial trend of the expected rainfall values (Feki 2010). Arnaud & Emery (2000) mentioned that the representation of a variable by a semi-variogram is a subjective operation: there is no ‘real’ underlying model. In practice, it is necessary to ensure that the model meets the main characteristics of the experimental semi-variogram, i.e., behavior at the origin, to the infinity, etc. Moreover, the adjustment of the model is not done using only the experimental semi-variogram but must integrate all available information on the phenomenon. Therefore, using such automatic methods of fitting as least square method, cross-validation, etc. is not justified as much as the phenomenon must be well-known spatially.

Spatial variation of interannual rainfall

Interannual rainfall semi-variograms are calculated for the 101 rain-gauges and fitted with power type model (Figure 3) which indicates the presence of a certain spatial trend. The model semi-variogram equation is: $$\gamma(h) = 572,000 + 22.9 \times h.$$
Spatial variation of seasonal rainfall

Seasonal rainfall semi-variograms are fitted with a power type model (Figure 4), indicating the presence of a spatial drift.

Spatial variation of monthly rainfall

Figure 5 shows the experimental and fitted sample semi-variograms of the monthly average rainfall for the four illustrative months: October, January, March, and June. The parameters of all the semi-variograms are given in Table 2. Except for the month of July, all the other months have a power type model of semi-variogram
which highlights the presence of a drift showing a large-scale trend.

**Nugget effect**

In theory, the semi-variogram value at the origin (0 lag distance) should be zero. If it is significantly different from zero for lags very close to zero, then this semi-variogram value is referred to as the nugget. The nugget represents variability at distances smaller than the typical sample spacing, including measurement error. In this study, the nugget effect values are important during the humid months (Figure 6) due to the great rainfall variability, with the largest value in December (34,200). Information changes rapidly in space during this season, which corresponds to the winter frontal disturbances coming from the north–west facing perpendicular channels of relief, and therefore large spatial rainfall variations (Slimani et al. 2007). Even if the network is relatively dense in the north, it turns out to be insufficient to account for certain micro-regionalization.

**Rainfall cartography**

In this study, we have assigned a deterministic status for the internal and external drift, and we have admitted that the semi-variograms of the main variable (rainfall) and the residuals are closer (Hudson & Wackernagel 1994; Bourennane & King 2003; Marcotte 2002). Based on these results, the simple semi-variogram parameters of the main variable (rainfall) are used and introduced into the kriging algorithms. The UK and the KED are applied for the estimation of the monthly, seasonal, and interannual average rainfalls. The four months of October, January, March, and June are used for illustration.

In general, the two methods show the same rainfall spatial repartition with some localized differences (Figures 7–9). The most important precipitation amounts concern the north-west region for the autumn, winter, and spring seasons, and the interannual map with the highest amount during winter (January) due to the fact that the isohyets are closer in that region than in the center and the south. During summer, the most important

![Figure 7](https://iwaponline.com/hr/article-pdf/48/2/514/366246/nh0480514.pdf)
Figure 8 | Seasonal rainfall maps. (Continued.)
Figure 8 | Continued.
Figure 9 | Monthly rainfall maps. (Continued.)
Figure 9 | Continued.
precipitations concern the high altitude of the dorsal. Furthermore, for the two methods, the direction of the monthly average rainfall isohyets changes with regions and seasons; in fact, in autumn (October), the isohyets that have a north-east/south-west direction in the north change to a north-west/south-east direction...
in the center and south of the country with a low gradient.

During the winter season (January), there is a visible important gradient in the north-west of the country, but in the center and the south there is no spatial structure. In spring, the isohyets keep the same direction, namely, the north-east/south-west, but with a more important gradient in the north than in the center. Finally, the interannual rainfall isohyets' disposition is a combination of the seasonal ones; therefore, we distinguish an important north-west/south-east gradient in the north-west and a less important gradient all over the center and the south.

Cross-validation

To assess the performance of each prediction method, cross-validation was used. This technique allows the comparison of the estimated and the actual values using the information available in our sample data set. The sample values are temporarily discarded from the sample data set; the value is then re-estimated using the remaining samples. The estimates are finally compared to the true values (Figure 10). There is a linear tendency between the actual and estimated values for all variables, and given by the two methods.

The two methods of interpolation show the same coefficient of determination between the estimated values and the actual values for autumn (Figure 11). From December to April, the KED method presents the largest value of the coefficient of correlation, while during summer the largest values are obtained by UK. In terms of the root mean square error (RMSE), the UK method provides the most accurate estimates for almost all months (Figure 12). The largest RMSE (and thus the most biased estimates) is for the winter (December and January).

Distribution of estimation uncertainty

One of the advantages of the kriging algorithms is the fact that they calculate the associated estimation uncertainty.
Figure 13 | Uncertainty maps of the seasonal rainfall. (Continued.)
Figure 13 | Continued.
This estimation uncertainty, which is quantified by calculating the standard deviation maps, increases as the observation network density decreases. The general pattern of the uncertainty repartition obtained by UK and KED, for seasonal rainfall (Figure 13), interannual rainfall (Figure 14), and monthly average rainfall (Figure 15), is the same with a random aspect of isohyets, these latter becoming more structured to the south and the west of the country. Except in summer (or dry months), the minimum uncertainty estimation values are minimum at a direct neighborhood of points measure; they become more important in the regions with sparse networks, essentially the West and the South of the country. This contrast is more clear for the monthly time scale than the interannual and seasonal ones. The standard deviation of uncertainty estimation is largest in winter, compared with other seasons, because of the perturbations occurring during this season. In general, the KED method provides slightly weaker maximal values of standard deviation of uncertainty estimation than the UK for all seasons. Furthermore, the maps show more detailed contours, that is, in regions with sparse networks, incorporating the elevation in the rainfall interpolation gives more accurate estimations than geographical coordinate introduction.

Localization of new rain-gauges

The locations of the additional rain-gauges can be optimized on the basis of the estimated variance. The simultaneous examination of the rainfall maps and standard deviation maps allowed the localization of new rain-gauges. Logically, regions where the standard deviation of the uncertainty estimation is high and where the rainfall gradient is important are the most concerned in increasing the density of stations. The comparison between the rainfall maps and the standard deviation uncertainty maps demonstrates that for the northern regions, the difference between the actual values and estimated values of rainfall is in the order of 10%, it increases to the center to about 25 to 30%, and it becomes more important to the south. The higher this percentage is, the worse is the estimation. The scenarios of increasing the monitoring network depend on the interpolation methods on one hand and the lag time on the other hand, but to ease the study a set of 17

![Figure 14](https://iwaponline.com/hr/article-pdf/48/2/514/366246/nh0480514.pdf)
Figure 15 | Uncertainty maps of the monthly average rainfall. (Continued.)
Figure 15 | Continued.
fictitious points (Figure 16) were chosen as new stations in the region with a sparse network. Then, the information gain at any point of the study area was assessed in terms of reduction of the kriging variance: this type of approach is, therefore, an ‘analysis of variance reduction’ (Cheng et al. 2008).

To assess the impact and the spatial range of the network strengthening on the estimation precision, the gain in accuracy is calculated. Let \( \sigma_0 \) be the standard deviation of the uncertainty estimation calculated over the 101 existing rain-gauges throughout Tunisia. If we implement a new rain-gauge at a point M, this standard deviation estimation will take a new value \( \sigma_M < \sigma_0 \). The percentage of the gain in accuracy can be defined as follows (Delhomme & Delfiner 1973):

\[
G(M) = \frac{\sigma_0 - \sigma_M}{\sigma_0} \times 100
\]

The mapping of this indicator (Figure 17) shows that the maximum gains obtained by UK and KED are in the direct vicinity of the new stations. They are generally more important in the center and the south than in the north of the country. Using the KED interpolation method (Figure 17(b)) shows a small gain of accuracy in the extreme north of the country for the interannual rainfall. That seems be generated by winter and spring gain amounts (Figure 17(d)). The gains in accuracy reach 80% in the summer and autumn season in the south, and they vary between 0 and 10% in the north in winter and spring (Figure 17). The range of influence of the added stations on the gain of precision is also different from one region to another (Figure 18); it is largest in the south and the center, in contrast with the north where the influence is more localized. The reason is principally the position of the added station versus the initial network stations’ positions. Finally, we notice that the two methods give slightly the same results, which means that the consideration of the altitude or the geographic coordinates in the kriging variance gave the same variance reduction.

The proposed approach to study the optimization of a network and decide about its capacity can be as follows:

- Calculate a semi-variogram that includes all events or any time lag of the studied field.
- Calculate and map the kriging standard deviation of this field.
- Identify areas of strongest local maxima of kriging standard deviation.
- Implement fictitious observation stations in these areas and re-calculate the kriging standard deviation.
- Calculate and map the gain in precision.
- Identify areas where the gain exceeds a given threshold, these areas are concerned by the implementation of new stations.
- Seek valuable sites in these areas to install actual station measures, responding to other informative and practical criteria such as the absence of obstacles, easy access, and availability of means of communication.

**CONCLUSION**

In this paper, we are interested in optimizing the design of the rain-gauge network in Tunisia based on reducing the variance of kriging. A first phase of statistic analysis, on mean monthly, seasonal, and interannual data, is followed by the computation of the experimental semi-variograms and the
Figure 17 | Gain maps. (Continued.)
semi-variogram model fitting. The analysis of these semi-variogram models has clearly highlighted some peculiar characteristics. In fact, all the months except July, seasons and interannual time scales have a power type model of semi-variogram showing a large-scale trend. UK and KED are applied to develop rainfall contour maps using geographical coordinates as internal drift and altitude as external drift, respectively. These methods are compared and it is shown that the method of KED performs better than the UK method in terms of RMSE. The plots of kriging standard deviation show contours of increasing magnitude where station density is minimal. A test implementation of some new stations is conducted throughout the country and the gain assessed. We have concluded that in the northern part of the country the system interpolation is not sensitive when adding additional information. Therefore, the existing network satisfactorily describes rainfall spatial variability. Indeed, the maximum obtained gain in precision exceeds 10%. In the center, the system is averagely sensitive when adding new stations, since the gain in precision varies between 20 and 40%. In the south, the gain accuracy can reach 80%. Consequently, the network is insufficient to
Figure 17 | Continued.
Figure 17 | Continued.
take into account the rainfall spatial variability, especially in autumn. A further planned improvement of the methodology consists of integrating the geographic coordinates and altitude as secondary variables altogether and by using other multivariate geostatistical methods such as regression kriging.

ACKNOWLEDGEMENTS

The authors thank the INM and the DGRE for providing rainfall data.

REFERENCES


Marcotte, D. 2012 *Geostatistics Course*. Polytechnique School of Montreal, Canada.


