Hydrology and
Earth System
Sciences

Open Access

Development of a method of robust rain gauge network optimization based on intensity-duration-frequency results

A. Chebbi[1]**, Z. K. Bargaoui**[1]**, and M. da Conceição Cunha**[2]

[1]Université Tunis El Manar, Ecole Nationale d'Ingénieurs de Tunis, BP 37 1002 Tunis, Tunisia
[2]Civil Engineering Department, University of Coimbra Polo II da Universidade-Pinhal de Marrocos, 3030-290 Coimbra, Portugal

Correspondence to: A. Chebbi (bachebbiaf@yahoo.com)

Abstract. Based on rainfall intensity-duration-frequency (IDF) curves, fitted in several locations of a given area, a robust optimization approach is proposed to identify the best locations to install new rain gauges. The advantage of robust optimization is that the resulting design solutions yield networks which behave acceptably under hydrological variability. Robust optimization can overcome the problem of selecting representative rainfall events when building the optimization process. This paper reports an original approach based on Montana IDF model parameters. The latter are assumed to be geostatistical variables, and their spatial interdependence is taken into account through the adoption of cross-variograms in the kriging process. The problem of optimally locating a fixed number of new monitoring stations based on an existing rain gauge network is addressed. The objective function is based on the mean spatial kriging variance and rainfall variogram structure using a variance-reduction method. Hydrological variability was taken into account by considering and implementing several return periods to define the robust objective function. Variance minimization is performed using a simulated annealing algorithm. In addition, knowledge of the time horizon is needed for the computation of the robust objective function. A short- and a long-term horizon were studied, and optimal networks are identified for each. The method developed is applied to north Tunisia (area = 21 000 km^2). Data inputs for the variogram analysis were IDF curves provided by the hydrological bureau and available for 14 tipping bucket type rain gauges. The recording period was from 1962 to 2001, depending on the station. The study concerns an imaginary network augmentation based on the network configuration in 1973, which is
a very significant year in Tunisia because there was an exceptional regional flood event in March 1973. This network consisted of 13 stations and did not meet World Meteorological Organization (WMO) recommendations for the minimum spatial density. Therefore, it is proposed to augment it by 25, 50, 100 and 160 % virtually, which is the rate that would meet WMO requirements. Results suggest that for a given augmentation robust networks remain stable overall for the two time horizons.

1 Introduction

Rain gauge monitoring networks are highly important to estimate precipitation and erosion, and to evaluate runoff. In fact, rainfall intensity is the most important variable in runoff and erosion impact prediction. More generally, rainfall network accuracy depends on precipitation variability as well as on the network size and design. Several optimization approaches have been proposed in the literature since the early work of Bras and Rodríguez-Iturbe (1976) and Delhomme (1978), who proposed a methodology of network design based on the minimization of the mean areal kriging error variance. The adoption of geostatistical methods for rainfall network sizing and augmentation was also performed by Pardo-Igúzquiza (1998). In Delhomme (1978), the optimal location of rain gauges was identified using a technique called the fictitious point method while in Pardo-Igúzquiza (1998) an automatic optimization technique (namely simulated annealing) was adopted. Barca et al. (2008) provided a methodology for assessing the

optimal location of new monitoring stations within an existing rain gauge monitoring network. The methodology used geostatistics and probabilistic techniques (simulated annealing) combined with GIS. A method composed of kriging and entropy that can determine the optimum number and spatial distribution of rain gauge stations in catchments was proposed in Chen et al. (2008). Chebbi et al. (2011) have considered mono-objective criteria assuming 1 h rainfall intensity interpolation and erosivity factor interpolation and using one single extreme rainfall event to conduct the analysis. Rainfall quantities retained in previous studies were mainly taken in a deterministic way. Effectively, a single rainfall pattern was selected for which the average kriging variance was minimized to achieve the best new rain gauge locations (Delhomme, 1978; Pardo-Igúzquiza, 1998; Chebbi et al., 2011). In the present study, it is aimed to find out new observation locations using a collection of rainfall patterns or rainfall auxiliary variables, each characterized by its probability of occurrence. Because robust optimization is an approach which can deal with the uncertainty in optimization problems by computing a solution that can cope with possible different scenarios (Mulvey et al., 1995; Bai et al., 1997; Beyer and Sendhoff, 2007), we claim that a robust network augmentation framework is proposed here.

Regarding water related problems, the literature contains a few applications of robust optimization techniques. Watkins and McKinney (1997) proposed two problems to illustrate the suitability of robust optimization in the resolution of water resources problems. The first is a problem of urban transfer of water vis-à-vis the availability of water and the need to consider water supply as random variable for decision making. The second problem relates to the management of the quality of subsoil waters, considering the uncertainty of the aquifer parameters. Ricciardi et al. (2007) considered a similar question in the context of aquifer remediation. Afonso and Cunha (2007) developed a robust model to design biological reactors and secondary settling tanks in wastewater treatment plants. Cunha and Sousa (2010) presented models for the robust design of water distribution networks to enable them to face the uncertainty of network working conditions under extreme events. Zeferino et al. (2012) have recently proposed a robust optimization model for the sitting and sizing of wastewater treatment plants at regional level that includes uncertainty issues associated with river flows. Accidents such as broken conduits or tanks and change in demand may affect how water distribution functions. Our study proposes to apply these approaches to decide on rain gauge network development. The problem of the best rain gauge location is addressed. North Tunisia is the study domain (area = 21 000 km^2). Rainfall intensity at a given location is considered as a random variable. Local intensity-duration-frequency curves are assumed to reflect the hydrological variability. Section 2 presents the method used in this paper. Section 3 presents the case study and the avail-

able data. Section 4 sets out the results obtained, and some concluding remarks are presented in Sect. 5.

2 Method

2.1 Definition of candidate stations

The main purpose of the rain gauge selection algorithm is to identify an optimal set of locations for a particular number of stations over the study area. The domain is in the Mediterranean area, and elevation ranges from 0 to 1281 m. According to WMO (1994, Sect. 20), the minimum recommended density is 1 station per 600 km^2 for Mediterranean plain areas in difficult conditions. This means, for instance, places where gauges are difficult to install and maintain, perhaps because of rugged topography or site inaccessibility. The initial network we consider was in operation at the time of the March 1973 flood event, which is why we took this as the point of departure of our problem. It was composed of 13 sparsely distributed stations, which was wholly inadequate to cover the rainfall variability over the study area. This study has a methodological character, and so various scenarios are simulated where the number of stations of the initial network is increased by 25 % (scenario 1), 50 % (scenario 2), 100 % (scenario 3) and 160 % (scenario 4). Moss and Tasker (1991) suggested that the number of candidate stations should be at least three times the number of the desired optimal stations. Accordingly, for scenarios 1 to 3, 40 candidate stations are assumed. In scenario 4, which achieves WMO-required density, the number of candidate stations is increased to 60. These 60 candidate stations, which contain the 40 candidate stations considered previously, are imaginary locations of rain gauges, equally distributed over the study area.

2.2 The IDF database

The search for "robust" solutions is an adequate method to solve the problem as it takes into account the hydrological uncertainty inherent to rainfall occurrence. Since we are interested in short duration rainfall, the maximum rainfall intensity for specified durations is the variable to be studied in this paper. To deal with hydrological risk inherent to rainfall data, we would need data on the maximum rainfall intensities recorded for several events. However, the problem is that we do not have this type of information for the study area. Thus, the adjustment parameters of the intensity-duration-frequency (IDF) curves (Koutsoyiannis et al., 1998) are proposed as alternative or auxiliary variables (parameters $a(T)$ and $b(T)$ of Eq. 1). An existing IDF study performed by DGRE-ST2i (2007) is taken as basis for this. The following times of reference were considered in DGRE-ST2i (2007): 5, 10, 15, 30, 45,..., 180 min. For stations having short observation periods (3 to 10 yr) without gaps, DGRE-ST2i (2007) selected thresholds (generally one threshold per time of reference), identified using various

tests, and adopted the intensities that were greater than the fixed threshold, to constitute the time series. The peak over threshold approach was adopted by DGRE-ST2i (2007) for the rain gauges characterized by a large number of gaps in the rainy months, even in the case where the number of observation years exceeded 10. For the rain gauges characterized by recordings without gaps and observed over long periods, DGRE-ST2i (2007) considered the M highest values observed in M years to achieve the statistical analysis.

The statistical study was carried out by DGRE-ST2i (2007) with Hydraccess software from IRD (Hydraccess, 2000). The series of annual maximum rainfall intensities for reference durations of 5, 10, 15, 20, 30, 45, 60, 120 and 180 min were adjusted by DGRE-ST2i (2007) by means of nine probability distribution functions: Gauss, Gumbel, Galton, Pearson III, Pearson V, Goodrich, Fréchet, WRC-USA and Escapes.

The Montana model, which predicts maximum rainfall intensity over duration t for the return period T as a power function of the duration (Eq. 1) (Burlando and Rosso, 1996), was adopted by DGRE-ST2i (2007). The estimated model parameters $a(T)$ and $b(T)$ are reported.

$$I(t, T) = a(T) * t^{-b(T)}, \tag{1}$$

where T is the return period (in years). Here, we adopt the hydrological risk definition where the risk is $p = 1/T$ (Bobée and Ashkar, 1993) so that we may consider that T reflects the hydrological risk, t is rainfall duration in minutes, and $a(T)$ and $b(T)$ are Montana IDF model parameters.

2.3 Geostatistical framework for IDF parameters

In this study $a(T)$ and $b(T)$ are taken as geostatistical variables (Matheron, 1965). In fact, it is assumed that it is possible to represent the $a(T)$ and $b(T)$ spatial structures through variogram functions. Furthermore, the analysis is made possible by the fact that the two parameters $a(T)$ and $b(T)$ are known at the same experimental points.

Further, we first test whether $a(T)$ and $b(T)$ are dependent or independent. For this, their experimental cross-variogram $\gamma_{a(T)b(T)}$ (Chilès and Delfiner, 1999) is examined (Eq. 2). The experimental cross-variogram is defined as the variance of the difference between two variables of different types or attributes at two locations.

$$\gamma_{a(T)b(T)}(h) = \frac{1}{2 * N(h)} \sum_{i=1}^{N(h)} [a(x_i + h) - a(x_i)]$$
$$[b(x_i + h) - b(x_i)] \tag{2}$$

where $N(h)$ represents the number of sample points separated by interdistance h, and x_i and $x_i + h$ are sampling locations separated by interdistance h.

For the cross-correlation analysis, it is recommended to adopt the co-dispersion coefficient graph $r_{a(T)b(T)(h)}$ (Matheron, 1965), which is linked to the cross-variogram and to the

direct variograms by

$$r_{a(T)b(T)}(h) = \gamma_{a(T)b(T)}(h) \Big/ \sqrt{\gamma_{a(T)a(T)}(h) * \gamma_{b(T)b(T)}(h)} \tag{3}$$

where $\gamma_{a(T)a(T)}$ and $\gamma_{b(T)b(T)}$ are the direct variograms respectively of $a(T)$ and $b(T)$.

Generally, the direct variogram function is a key tool to quantify the variability associated with the regionalized variable, Z (the variable Z is set as $a(T)$ or $b(T)$). The experimental semivariogram, $\gamma(h)$, is calculated from the data as a function of the point separation, h, and is given by

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i + h) - Z(x_i)]^2, \tag{4}$$

where $N(h)$ is the number of sample points separated by h, x_i and $x_i + h$ are sampling locations separated by a distance h, and $Z(x_i)$ and $Z(x_i + h)$ are values of the variable Z measured at the corresponding locations.

The sample variogram is fitted to a variogram model. For a stationary regionalized variable, the variogram is characterized by three main parameters: range, sill, and nugget. "Range" is the distance at which measurements cease to be correlated with each other. "Sill" is the variogram value at and beyond the range distance. The "nugget" effect is the random component of the digital values, graphically expressed by the discontinuity of the variogram at the origin. A parametric approach is used to derive the variogram model. Without loss of generality, further developments are given for the example of the spherical model in Eq. (5), which is a bounded variogram:

$$\gamma_{\delta}(h) = s(\delta) * \left[1.5 * (h/r(\delta)) - 0.5 * (h/r(\delta))^3\right]. \tag{5}$$

The sill $s(\delta)$ represents the highest variance for a large data point distance, and, for the spherical model, the range $r(\delta)$ refers to the distance over which the data are correlated.

The crossed and direct variograms are estimated from the pairs $(a(T), b(T))$ estimated at each observed location for the various durations ($t = 5, 10, 15, 20, 30, 45, 60, 120$ and 180 min).

It is assumed that if the graph $r_{a(T)b(T)}(h)$ is constant, it may be concluded that the parameters $a(T)$ and $b(T)$ are dependent (Chilès et al., 1991).

Furthermore, network optimization is achieved by adopting an objective function to be maximized (or minimized). The quality of $a(T)$ spatial interpolation is selected as a target, and kriging is adopted as the interpolation method. Effectively, advantage is taken of the fact that kriging is accompanied by the estimation of the variance of estimation error, which is an indicator of interpolation accuracy. Also, an advantage of kriging is that kriging variance error can be simulated using imaginary networks (with no observed data at the krigged location). We therefore used kriging methods to define the objective function. Hence, the minimization of the
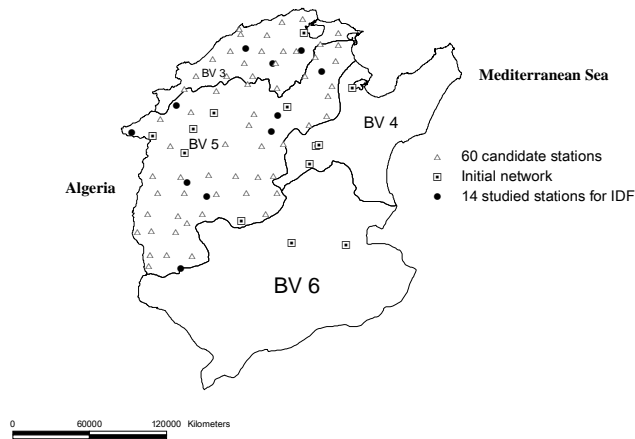
Fig. 1. Location of rain gauges in the study area (Medjerda basin BV 5; northern coast basin BV 3; and surrounding basins (Meliane basin BV 4 and central Tunisia basin BV 6)).

Table 1. Studied rain gauges by DGRE-ST2i (2007).

Rain gauge name	Rain gauge number	Basin	Time series size	Begining of recording
Ghardimaou	1	BV 5	26 yr	29 Dec 1973
Zouarine gare	2	BV 5	33 yr	30 Sep 1968
Aïn Taga	3	BV 5	32 yr	9 May 1964
Aïn Beya Fernana	4	BV 5	17 yr	1 Jan 1983
Haïdra Poste Douanes	5	BV 5	17 yr	5 Mar 1984
Izid Barrage	6	BV 5	10 yr	28 Jul 1973
Joumine Antra	7	BV 3	28 yr	21 Aug 1963
Mellègue K13	8	BV 5	17 yr	15 Dec 1976
Oued Tine cassis	9	BV 3	20 yr	12 Jul 1968
Sarrat Pont Route	10	BV 5	18 yr	21 Oct 1982
Sejnène	11	BV 3	15 yr	19 Sep 1962
Siliana Laouej	12	BV 5	14 yr	5 Feb 1974
Slouguia	13	BV 5	15 yr	3 Jan 1976
Sraya Ecole	14	BV 5	26 yr	16 Dec 1975

average (spatial mean over the study domain) kriging error of $a(T)$ is the specific approach proposed here. Six return periods ($T = 2, 5, 10, 20, 50, 100$ yr) covering a broad panoply of risk situations are considered.

As $a(T)$ and $b(T)$ are not independent, it is proposed to develop the kriging estimate of $a(T)$ using $b(T)$ as information. External drift kriging (EDK) is a suitable method to achieve this goal. In fact, in the case of a sparse network, kriging with external drift seems more appropriate than ordinary kriging. EDK requires knowledge of the values of $b(T)$ at the locations where $a(T)$ is to be interpolated. This is achieved by first using an ordinary kriging approach to $b(T)$. Therefore, a map of $b(T)$ is produced that is then adopted to interpolate $a(T)$. The kriging systems are set out in Appendix A.

2.4 The objective function

The decision model presented here is built within the framework of robust optimization and is inspired by the case studies reported in Mulvey et al. (1995). The objective function is written using the concept of regret by considering a quadratic term expressing the difference, for each scenario, between the values of the standardized mean spatial kriging variance achieved by the solution to be implemented and by the optimal solution for the scenario. This means that the optimal solution for the model proposed will be solution-robust (Laguna, 1998). As such, the optimal solution obtained will be "close" to the optimum for any of the realized scenarios ensuring the optimality robustness. The robust optimization method adopted requires knowledge of a time horizon. For a fixed hydrologic risk $p = 1/T$, where T is the return period, we express the probability of an overrun of the event of return period T during time horizon of duration N. It corresponds

to

$$u(T) = 1 - (1 - p)^N, \tag{6}$$

where N is the number of years in the horizon. Two time horizons are successively considered in the present study: the short term ($N = 5$ yr) and the long term ($N = 30$ yr).

$u(T)$ is further scaled by dividing it by the sum of $u(T)$ over the various return periods.

$$\omega(T) = u(T) \Big/ \sum_{i=1}^{i=NT} u(T_i), \tag{7}$$

where NT is the number of return periods; this means the number of scenarios considered in the study.

To evaluate the mean spatial kriging error variance over the study domain, a grid mesh with a resolution of 4 km was used. The optimization problem consists of minimizing the objective function expressed by

$$\text{Min} \sum_{i=1}^{NT} \omega(T = T_i) * (S(T = T_i) - S_{\text{ref}}(T = T_i))^2, \tag{8}$$

where $\sum_{i=1}^{NT} \omega(T = T_i) = 1$ with $\omega(T = T_i)$ as indicated in (Eq. 7), with $S_{\text{ref}}(T = T_i)$ being the value of the standardized mean spatial kriging variance obtained for every return period T_i independently of the other return periods. It is taken as reference.

In addition, standardization of the mean spatial kriging variance is obtained by using the interquartile range of $a(T)$ kriging error variance map:

$$S(T = T_i) = \left(\sum_{i=1}^{n} \left(\sigma_{i(a(T=T_i))} \right)^2 \Big/ n \right) \Big/ F(T = T_i), \tag{9}$$

(a) case 1 : T = 2 years

(b) case 2 : T = 5 years

(c) case 3 : T = 10 years

(d) case 4 : T = 20 years

(e) case 5 : T = 50 years
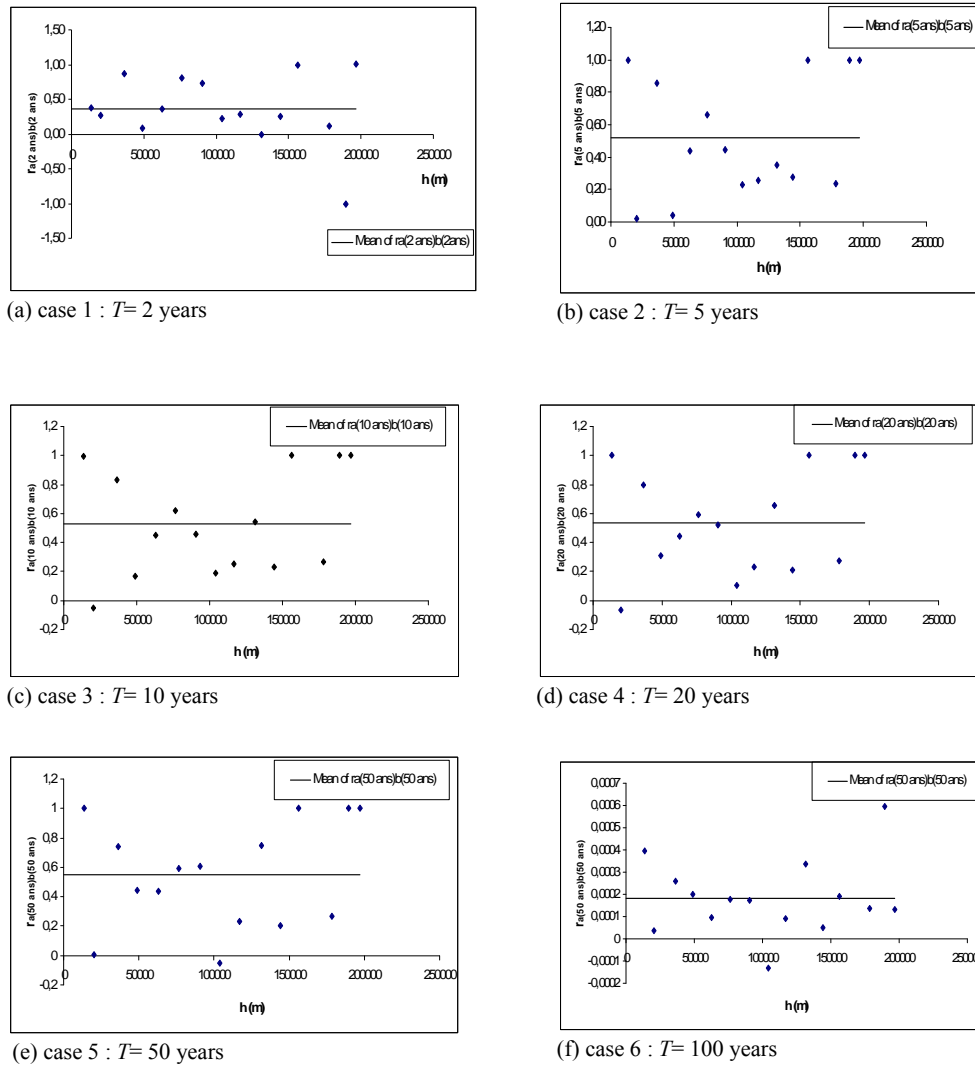

(f) case 6 : T = 100 years

Fig. 2. Fluctuation of $r_{a(T)b(T)}(h)$.

where $(\sigma^2_{i(a(T=T_i))})$ is the variance of kriging errors of $a(T = T_i)$ at the computing node i depending on locations of stations, and n is the number of grid nodes.

$$F(T = T_i) = \left(\sigma^2_{75\%(a(T=Ti))} - \sigma^2_{25\%(a(T=Ti))} \right) \qquad (10)$$

$\sigma^2_{75\%(a(T=T_i))}$ is the 75th percentile of the pattern of the variance of kriging errors of $a(T = T_i)$.

$\sigma^2_{25\%(a(T=T_i))}$ is the 25th percentile of the pattern of the variance of kriging errors of $a(T = T_i)$.

This objective function is subjected to domain constraints expressed by the set of possible locations for the stations (as such the solution space is defined).

In all cases (for the two horizons with the four different scenarios of network augmentation (+25 %; +50 %; +100 %; +160 %)), the minimization problem is solved using a simulated annealing algorithm (Kirkpatrick et al., 1983).

3 Case study and data

The study area is composed of two watersheds: the north coast watershed (BV 3) and the Medjerda watershed (BV 5) in Tunisia. The study area is characterized by a sub-humid to semi-arid climate and covers more than 20 000 km^2. Figure 1 shows the study area together with the reference rain gauges of the DGRE-ST2i study composed of 14 stations. It also shows the initial network consisting of the 13 stations functioning in March 1973 during the extreme flood event. Candidate locations for composing the robust network are also marked. Table 1 presents the rain gauges from the DGRE-ST2i study. The Montana IDF model parameters estimated by DGRE-ST2i (2007) are presented in Table 2 for the 14 stations. The size of the time series analysed in DGRE-ST2i ranges from 10 to 33 yr (Table 1). The observations periods range from 1962 to 2001, depending on the station.

Table 2. Parameters $a(T)$ and $b(T)$ for the 14 stations studied for IDF (DGRE-ST2i, 2007).

Rain gauge number	Return period 2 yr		Return period 5 yr		Return period 10 yr		Return period 20 yr		Return period 50 yr		Return period 100 yr	
	$a(T)$	$b(T)$	$a(T)$	$b(T)$	$a(T)$	$b(T)$	$a(T)$	$b(T)$	$a(T)$	$b(T)$	$a(T)$	$b(T)$
1	231.6	0.685	289.9	0.69	324.4	0.688	356.9	0.686	401.3	0.685	435.7	0.685
2	240	0.66	303	0.64	352	0.63	404	0.63	479	0.62	542.0	0.62
3	437.1	0.538	541.8	0.53	614	0.524	680	0.517	759.9	0.507	812.4	0.498
4	368.5	0.682	528.6	0.694	629.6	0.683	715	0.665	804	0.635	856.4	0.61
5	321.4	0.745	406	0.741	456.9	0.734	503.4	0.726	556.9	0.713	594.4	0.704
6	157.7	0.652	307.7	0.73	465.9	0.776	651.5	0.81	920.4	0.84	1142.3	0.857
7	179	0.549	220	0.511	249.5	0.49	277.7	0.473	313.8	0.455	341.8	0.445
8	410.4	0.722	553	0.68	668.7	0.654	793.3	0.635	965.1	0.616	1103.9	0.606
9	144.5	0.626	206.3	0.637	241.9	0.624	286.8	0.618	345.9	0.605	392.4	0.594
10	431.5	0.758	688.3	0.789	848.3	0.796	992.5	0.798	1171.5	0.797	1317.5	0.798
11	304.8	0.669	349.2	0.663	385.4	0.66	429.2	0.661	499.2	0.666	559.9	0.671
12	254.2	0.635	342.6	0.675	406	0.692	461.1	0.698	512.4	0.689	530.8	0.67
13	275.1	0.695	343.8	0.69	387.7	0.682	427.6	0.673	479.7	0.663	519.8	0.657
14	162.1	0.592	200.1	0.592	237.2	0.596	284.7	0.604	367.1	0.619	448.1	0,633

Table 3. Adjusted sills and ranges for $a(T)$ and $b(T)$ assuming spherical variogram models.

Return period T (yr)	Sill b	Range b (km)	Sill a	Range a (km)
2	0.004	30	15 000	40
5	0.008	40	32 000	50
10	0.010	40	45 000	45
20	0.014	40	50 000	50
50	0.016	40	110 000	50
100	0.018	40	110 000	50

Table 4. Values of $u(T)$ and $\omega(T)$.

T (yr)	Horizon = 5 yr		Horizon = 30 yr	
	$u(T)$	$\omega(T)$	$u(T)$	$\omega(T)$
2	0.97	0.56	1	0.41
5	0.67	0.39	1	0.41
50	0.10	0.05	0.45	0.18

Table 5. Values of $\mathrm{OF_{ref}}$ (T) objective function obtained for interpolation of $a(T)$.

Return period T	2 yr	5 yr	50 yr
25 % (scenario 1: 3 new stations)	2.13	1.69	1.67
50 % (scenario 2: 6 new stations)	1.52	1.34	1.35
100 % (scenario 3: 13 new stations)	1.30	1.05	1.06
Normative density of the WMO (160 %) (scenario 4: 21 new stations)	1.03	0.89	0.89

3.1 Results

3.2 Dependence of the parameters $a(T)$ and $b(T)$

The co-dispersion coefficient graph $r_{a(T)b(T)}(h)$ is plotted in Fig. 2. An erratic fluctuation of $r_{a(T)b(T)(h)}$ around a fixed value can clearly be seen. Therefore, $r_{a(T)b(T)}(h)$ can be assumed as nearly constant for all T values, allowing the supposition that $a(T)$ and $b(T)$ are dependent variables. This authorizes the kriging of $a(T)$ by taking $b(T)$ as external drift.

3.3 Structural analysis of the parameters $a(T)$ and $b(T)$

Spherical variogram models are adjusted to sample variograms. The ranges and sills were identified manually, and an attempt was made to take account of a good approximation of the first points of the sample variograms (short interdistances) as well as a good approximation of high inter-

distances. Table 3 reports the adjusted sill and range for every return period. Adjusted ranges of $a(T)$ and $b(T)$ extend from 30 to 50 km. Sample and adjusted variograms of $a(T)$ and $b(T)$ are plotted in Fig. 3. Instead of using all the return periods previously studied, it is proposed to adopt representative return periods. In effect, the analysis highlights three groups of return periods where the variograms are found to be quite similar: group 1 including only $\{T = 2\,\mathrm{yr}\}$; group 2 including $\{T = 5, 10, 20\,\mathrm{yr}\}$ and group 3$\{T = 50, 100\,\mathrm{yr}\}$. Hence, $T_1 = 2\,\mathrm{yr}$ (normal situation), $T_2 = 5\,\mathrm{yr}$ (small risk) and $T_3 = 50\,\mathrm{yr}$ (high risk) were selected as representative of each group.

(a) case 1 : Variogram of *b(2 years)*

b) case 2 : Variogram of *a(2 years)*

(c) case 3 : Variogram of *b(5 years)*

d) case 4 : Variogram of *a(5 years)*

(e) case 5 : Variogram of *b(10 years)*

f) case 6 : Variogram of *a(10 years)*

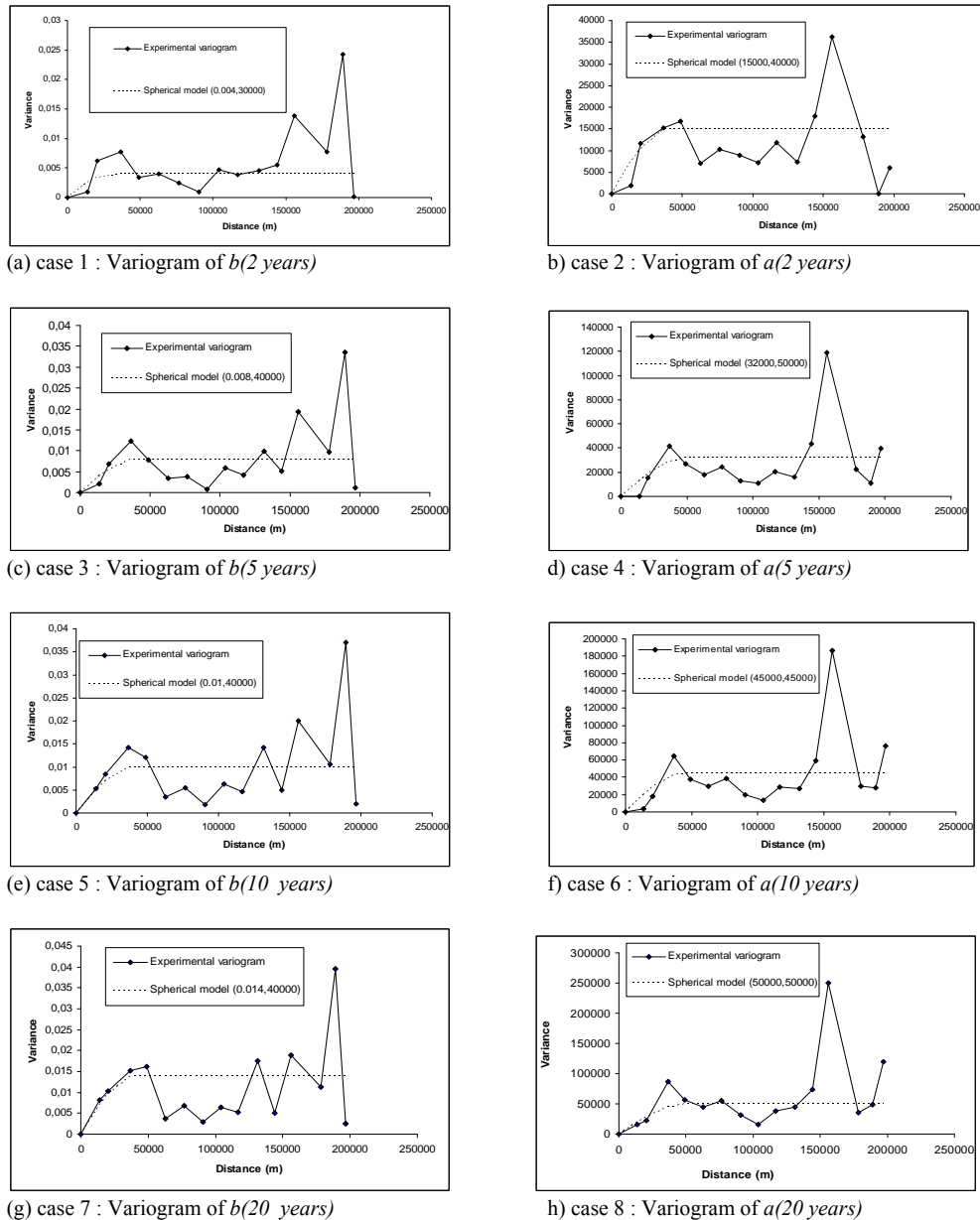(g) case 7 : Variogram of *b(20 years)*

h) case 8 : Variogram of *a(20 years)*

Fig. 3. Variograms of $a(T)$ and $b(T)$.

3.4 Comparison of the resulting robust networks

We first calculated the probability of overrun of the event $u(T)$ during the time horizon (Eq. 6) and thus the associated weight $\omega(T)$ (Eq. 7). Table 4 summarizes the values of $u(T)$ and $\omega(T)$ for each return period for the two horizons. In the case of the short-term time horizon, we find that $T = 50$ yr is nearly neglected, and there is more focus on $T = 2$ yr in the weighting system. For the long-term time horizon, the method assigns equal weights to normal and moderate risk (2 and 5 yr) while more weight is assumed for high risk situations ($T = 50$ yr). Therefore, these findings are consistent

with the intuitive point of view. Table 5 reports the values of the standardized variances $OF_{ref}(T)$, which obviously decreased as the network size increased (Table 5). It was also found that the most improvement is obtained when one splits from scenario 1 (three new stations to implement) to scenario 2 (six new stations to implement).

The comparison of the resulting robust networks obtained in turn for the two time horizons (short term and long term) is reported in Table 6, where the locations which are different for the two horizons are in italic. It shows that optimal networks are similar for scenario 1. This is an important result: the locations of the 3 new stations chosen from 40 candidates

Table 6. Robust solutions. The locations which are different for the two horizons are in italics.

Increasing the existing network by	horizon = 5 yr	horizon = 30 yr
Normative density of the WMO (160 %) (scenario 4: 21 new stations)	{22, *24*, 25, 26, 27, 29, 30, 31, *32*, 34, 40, *45*, *46*, 47, *48*, *49*, 50, 51, 53, 54, *55*}	{22, 25, 26, 27, *28*, 29, 30, 31, *33*, 34, *35*, *36*, 40, 47, 50, 51, *52*, 53, 54, *58*, *60*}
100 % (scenario 3: 13 new stations)	{12, *13*, 16, 18, 19, 20, 21, 22, 25, 27, *28*, 31, 35}	{12, 16, 18, 19, 20, 21, 22, 25, 27, 31, *32*, 35, *40*)}
50 % (scenario 2: 6 new stations)	{15, 18, 21, 25, 27, *34*}	{15, 18, *20*, 21, 25, 27}
25 % (scenario 1: 3 new stations)	{ 25, 27, 28}	{ 25, 27, 28}
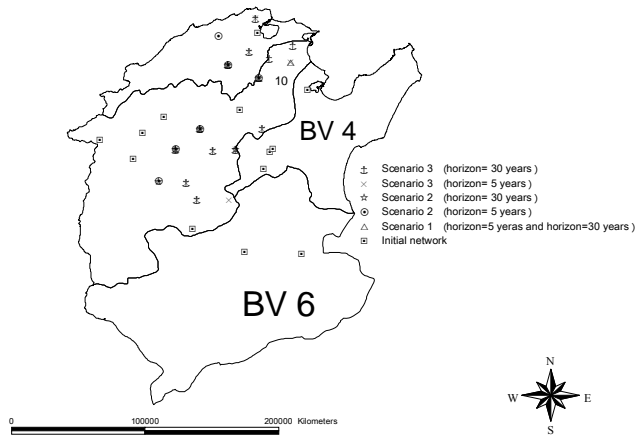
Fig. 4. Configuration of optimal networks.

are independent of the time horizon. This kind of network reinforcing is needed, without doubt. For scenario 2, which consists of adding six new stations chosen from 40 candidates, the two robust networks differ only by one station, for both time horizons. This also indicates good accuracy of the result. Five new stations can be added, and they are equally representative for the short- and long-term perspectives. For scenario 3, which requires 13 new locations to be found (out of 40 candidates), the two robust networks obtained in turn for the short- and long-term horizons differ only by 2 stations. Hence, as many as 11 new locations are common to the two time horizon perspectives, which is a very encouraging result from a decision making point of view. However, using 60 candidates for scenario 4, the short and long-term visions differ by 7 stations out of 21 (Fig. 4), which may be seen as problematic by a network manager.

In a previous paper (Chebbi et al., 2011), mono-objective criteria have been considered assuming 1 h rainfall intensity interpolation and erosivity factor interpolation and using one single extreme rainfall event to conduct the analysis. The comparison of previous results with the present study highlights that the mean spatial kriging variance in the case of the mono-objective criterion is lower than or equal to that obtained in the case of the robust optimization. Nevertheless, the essential advantage of the robust optimization lies in the fact that it allows overcoming the problem of using one single rainfall event and yields networks which work "adequately",

when considering various extreme events with different return periods.

4 Conclusions

The robust optimization approaches are recommended in cases where the variables of interest are uncertain . The hydrological risk is considered in the present study, which aims to find new observation locations for rain gauges for recording short duration rainfall. The novel approach consists of considering IDF curve parameters to locate the best sites for installing imaginary new rain gauges. The method assumes a time horizon to minimize an objective function-based IDF parameter $a(T)$ of the Montana model, considering the weighting of various return periods T. Kriging interpolation using the variance-reduction method was applied to build the objective function using the variance error of $a(T)$ estimation. The weighted mean spatial error variance was considered.

The region of north Tunisia, which has a sub-humid to semi-arid climate, was used to develop the methodology. The comparison of the resulting robust networks considered three return periods ($T_1 = 2$ yr, $T_2 = 5$ yr and $T_3 = 50$ yr), representing normal to high risk situations and two different time horizons (short term = 5 yr, and long term = 30 yr). It is suggested that robust networks are quite similar to each other when selecting 3, 6 and 13 locations from 40 candidates to reinforce an initial network of 13 stations covering an area of 21 000 km². The conclusions are quite different when selecting 21 new stations: only 14 out of 21 stations, assuming 60 candidate locations, are common to the two time horizons. Therefore, the conclusion that may be drawn is that when the size of the new network is augmented, it is more difficult to obtain a unique robust solution. Further research topics aim to develop entropy approaches to define the robust objective function.

Appendix A

Kriging systems

Once a proper variogram model is chosen, kriging is applied to the entire area of study to estimate the variable values at unsampled points, using the data from the surrounding sampled area. The kriging estimator is expressed as follows:

$$Z^*(x_0) = \sum_{i=1}^{N} \lambda_i Z(x_i), \tag{A1}$$

where $Z^*(x_0)$ is an estimated value of Z at location x_0, λ_i is the weight assigned to the observation at the location x_i, and N is the number of observations within the search neighbourhood.

The λ_i's are kriging weights which are estimated as the solution of the ordinary kriging system (Eq. 2):

$$\begin{cases} \sum_{i=1}^{N_{nb}} \lambda_j \gamma (x_j - x_i) + \mu' \\ \quad = \gamma (x_j - x_0) \ \text{j} = 1, ..., \text{N}_{nb} \quad , \\ \sum_{i=1}^{N_{nb}} \lambda_i = 1 \end{cases} \tag{A2}$$

where μ' is a Lagrange parameter accounting for the constraints on the weights.

The kriging variance for ordinary kriging is obtained by

$$\sigma_0^2 = \gamma(0) - \sum_{i=1}^{N_{nb}} \lambda_i \gamma (x_i - x_0) - \mu'. \tag{A3}$$

In the case of a sparse network, kriging with external drift seems more appropriate than ordinary kriging. Kriging with external drift represents the kriging estimates, $Z*(x)$, as a sum of a trend component $m(x) = E[Z(x)]$ and a residual $R(x)$ (Bardossy and Lehmann, 1998):

$$Z^*(x) = m(x) + R(x). \tag{A4}$$

The trend component is then replaced by

$$E[Z(x)|Y(x)] = m(x) = a_0 + a_1 Y(x), \tag{A5}$$

where a_0 and a_1 are unknown constants. The linear estimator (Eq. 1) should be unbiased for any a_0 and a_1 values.

The external drift kriging variance is obtained by

$$\sigma_0^2(x) = \gamma(0) - \sum_{i=1}^{N} \lambda_i \gamma (x_i - x) - \mu_1 - \mu_2 Y(x), \tag{A6}$$

where μ_1 and μ_2 are two Lagrange parameters accounting for the constraints on the weights. Therefore, Y has to be known at locations x_i as well as at the target location x_0.

References

Afonso, P. and Cunha, M. C.: Robust optimal design of activated sludge bioreactors, J. Environ. Eng., 133, 44–52, 2007.

Bai, D., Carpenter, T., and Mulvey, J.: Making a case for robust optimization models, Manage. Sci., 43, 895–907, 1997.

Barca, E., Passarella, G., and Uricchio, V.: Optimal extension of the rain gauge monitoring network of the Apulian regional consortium for crop protection, Environ. monitor. Assess., 145, 375–386, 2008.

Bardossy, A. and Lehmann, W.: Spatial distribution of soil moisture in a small catchment. Part 1: Geostatistical analysis, J. Hydrol. (Amsterdam), 206, 1–15, 1998.

Beyer, H.-G. and Sendhoff, B.: Robust optimization – a comprehensive survey, Computer Methods in Applied Mechanics and Engineering, 196, 3190–3218, 2007.

Bobée, B. and Ashkar, F.: The gamma family and derived distributions applied in hydrology, Water Resources Publications, Littleton, Colorado, 203 pp., 1991.

Bras, R. L. and Rodríguez-Iturbe, I.: Network design for the estimation of areal mean of rainfall events, Water Resour. Res., 12, 1185–1195, 1976.

Burlando, P. and Rosso, R.: Scaling and muitiscaling models of depth-duration-frequency curves for storm precipitation, J. Hydrol., 187, 45–64, doi:10.1016/S0022-1694(96)03086-7, Fractals, scaling and nonlinear variability in hydrology, 1996.

Chebbi, A., Kebaili Bargaoui, Z., and Da Conceiçao Cunha, M.: Optimal extension of rain gauge monitoring network for rainfall intensity and erosivity index interpolation, J. Hydrol. Eng., 16, 665–676, 2011.

Chen, Y.-C., Wei, C., and Yeh, H.-C.: Rainfall network design using kriging and entropy, Hydrol. Process, 22, 340–346, 2008.

Chilès, J. P. and Delfiner, P.: Geostatistics, Wiley Series in Probability and Statistics: Applied Probability and Statistics, John Wiley & Sons Inc., New York, doi:10.1002/780470316993, Modeling spatial uncertainty, A Wiley-Interscience Publication, 1999.

Chilès, J. P., Gable, R., and Morin, R. H.: Contribution de la géostatistique à l'étude de la thermicité des rides océaniques du Pacifique. Cahiers de Géostatistiques, Fascicule 1, Compte-rendu des Journées de Géostatistique, 6–7 Juni 1991, Fontainebleau, 51–61, 1991.

Cunha, M. C. and Sousa, J. J. O.: Robust Design of Water Distribution Networks for a Proactive Risk Management, J. Water Resour. Plng. and Mgmt., 136, 227–236, 2010.

Delhomme, J. P.: Kriging in the hydrosciences, Adv. Water Resour., 1, 51–266, 1978.

DGRE-ST2i groupe STUDI: Optimisation des réseaux de suivi des ressources en eau en Tunisie: Etude de la pluviographie, Direction Générale des Ressources en Eau, 2007.

Hydraccess: IRD, available at: http://www.mpl.ird.fr/hybam/outils/ ha4_fr_dn.php (last access: 2007), 2000.

Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P.: Optimization by Simulated Annealing, Science, 220, 671–680, 1983.

Koutsoyiannis, D., Kozonis, D., and Manetas, A.: A mathematical framework for studying rainfall intensity-duration-frequency relationships, J. Hydrol., 206, 118–135, 1998.

Laguna, M.: Applying robust optimization to capacity expansion of one location in telecommunications with demand uncertainty, Manage. Sci., 44, S101–S110, 1998.

Matheron, G.: Les Variables régionalisées et leur estimation, Masson, Paris, 305 pp., 1965.

Moss, M. E. and Tasker, G. D.: An intercomparison of hydrological network-design technologies, Hydrol. Sci. J., 36, 209–221, 1991.

Mulvey, J. M., Vanderbei, R. J., and Zenios, S. A.: Robust optimization of large-scale systems, Operations Res., 43, 264–281, 1995.

Pardo-Iguzquiza, E.: Optimal selection of number and location of rainfall gauges for areal rainfall estimation using geostatistics and Simulated Annealing, J. Hydrol., 210, 206–220, 1998.

Ricciardi, K. L., Pinder, G. F., and Karatzas, G. P.: Efficient groundwater remediation system design subject to uncertainty using robust optimization, J. Water Resour. Plan. Manage., 133, 253–263, 2007.

Watkins, D. W. and McKinney, D. C.: Finding robust solutions to water resources problems, J. Water Resour. Plan. Manage., 123, 49–58, 1997.

World Meteorological Organization (WMO): Guide to hydrological practices, Data acquisition and processing, analysis forecasting and other applications, Design and evaluation of hydrological networks, WMO, 168, Geneva, Switzerland, 1994.

Zeferino, J., Cunha, M. C., and Antunes, A. P.: Robust Optimization Approach to Regional Wastewater System Planning, J. Environ. Manage., 109, 113–12, 2012.