European Water 3/4: 9-15, 2003.

Network design for drought monitoring by geostatistical techniques

B. Bonaccorso, A. Cancelliere and G. Rossi
Department of Civil and Environmental Engineering, University of Catania, Italy

Abstract: Monitoring of hydrometeorological events characterized by high variability in time and space requires the optimal selection of the number and location of the gauging stations within the area under study. A renewed interest about this topic has risen because of the increasing diffusion of automatic data acquisition networks, which forces to review the design criteria of traditional networks in order to reduce installation and management costs of the new gauges, as well as to better monitor in real time extreme hydrometeorological events, such as floods and droughts. In the present work, a methodology has been developed to assess the network capability to properly describe drought events, by verifying that the Root Mean Square Error (RMSE) of the Standardized Precipitation Index (SPI) in locations without gauging stations is lower than an acceptable value. The proposed approach, based on the use of kriging technique, has been applied to the automatic data acquisition network of rainfall stations managed by the Regional Hydrographic Service of Sicily. The results of the analysis allow to determine and compare the RMSE related to different network configurations in order to select a suitable network of stations to be adopted for drought monitoring in Sicily region.

Key words: kriging, SPI, hydrometeorological network, drought

1. INTRODUCTION

Automatic data acquisition networks are powerful tools able to help decision makers in extreme events risk management by means of real time data measurement, transfer and processing. The increasing diffusion of automatic networks imposes design criteria which take into account new requirements for real time monitoring of hydrometeorological phenomena. One of the most widely used criteria for network design consists in minimizing estimation error of the hydrometeorological variable under study in points where direct measurements are not available, based on the use of geostatistical techniques (Bastin et al.,1984; Bogardi and Bardossy, 1985; Bacchi, 1996; Pardo-Iguzquiza, 1998). Ideally one should be able to keep such an error below a fixed limit in the whole area, provided that the number and location of stations is adequate. Estimation error can be assessed by means of the theory of *optimal linear estimators* applied to *regional variables* (Matheron, 1965), which allows to express such an error as a function of the stations location within the area of interest.

According to this approach, in the present paper a methodology for selecting the optimal number of stations to be adopted within a drought monitoring system in Sicily has been developed. Such a methodology has been applied to Sicily by selecting the stations among the ones integrated in the automatic data acquisition network of the Regional Hydrographic Service (STIR). Then, different configurations of the drought monitoring system have been compared on the basis of the Root Mean Square Error (RMSE) of the SPI index (McKee et al., 1993), computed in points where no measure is available. The RMSE has been determined by applying the *kriging technique*, which assumes spatial stationarity for the variable under investigation. This condition is well satisfied because of the standardized nature of SPI.

The paper is organized as follows: after a brief review on SPI index, the proposed methodology is described in detail and an overview about kriging technique is given. Finally, the results of the application to the STIR hydrometeorological network is reported.

2. METHODOLOGY

2.1 Drought monitoring by SPI index

The SPI is a meteorological index originally introduced by McKee et al. (1993) for defining and monitoring drought, which takes into account different time scales at which drought phenomenon affects water uses. The index is based on an equiprobability transformation between the cumulative values of past monthly precipitation over *k* months and a standardized normally distributed value. In practice, computation of the index requires fitting a probability distribution to cumulative monthly precipitation series (e.g. k= 3, 6, 12, 24 months), computing the non-exceedence probability corresponding to such cumulative values and assuming the normal standardized quantile corresponding to such probability as the SPI. McKee et al. (1993) assumed cumulative precipitation gamma distributed and used maximum likelihood method to estimate the parameters of the distribution. Although McKee et al. (1993) originally proposed a classification restricted only to drought periods, it became customary to use the index to classify wet periods as well. Table I reports the climatic classification according to the SPI index provided by National Drought Mitigation Center (NDMC, http://www.ndmc.unl.edu).

Table 1. Wet and drought period classification according to the SPI index

Index value	Class
2.00 ≤ SPI	Extremely wet
1.50 ≤ SPI < 2.00	Very wet
1.00 ≤ SPI < 1.50	Moderately wet
-1.00 ≤ SPI < 1.00	Near normal
-1.50 ≤ SPI < -1.00	Moderate drought
-2.00 ≤ SPI < -1.50	Severe drought
SPI < -2.00	Extreme drought

The dimensionless and standardized nature of the index allows to compare droughts among regions with different climates, as well as droughts occurring during different seasons of the year.

2.2 Assessment of estimation error of SPI index

Let $Z(\underline{x})$ be a value of a stochastic process $Z(x)$, representing the hydrometeorological variable under study, in a generic point $\underline{x}=\{x,y\}$. According to the theory proposed by Matheron (1965), $Z(x)$ is defined as *regional variable* and its probabilistic structure is assumed to be spatially stationary. Under the hypothesis of second order stationarity and isotropy, the following equations hold (Delhomme, 1978):

$$E\left[Z(\underline{x})\right] = m = constant \tag{1.a}$$

$$Var\left[Z(\underline{x})\right] = \sigma^2 = constant \tag{1.b}$$

$$Cov\left[Z(\underline{x}+\underline{h}), Z(\underline{x})\right] = Cov(h) \tag{1.c}$$

where h is the modulus of the generic vector $\underline{h}$, which is equal to the distance between points $\underline{x}+\underline{h}$ and $\underline{x}$. The function $Cov(h)$ describes the spatial covariance structure of the hydrometeorological phenomenon and is generally obtained by interpolation of the so called experimental covariogram, i.e. the plot of the sample covariances between observations in several couples of stations versus the distance between the stations. Note that the SPI index can be reasonably considered as a second order stationary process, due to its standardized nature. Further, since the variance σ^2 of the index is equal to 1, in this case $Cov(h)$ corresponds to the correlation coefficient $\rho(h)$. The value of $Z(.)$ in a point $\underline{x_0}$ can be determined as a linear combination of the values observed in N neighbouring stations:

$$Z^*(\underline{x_0}) = \sum_{i}^{N} \lambda_i Z(\underline{x_i}) \tag{2}$$

where the coefficients λ_i are called weights. By using eq. (2) and eqs. (1.a), (1.b) and (1.c), it can be easily shown that the variance of the estimation error in $\underline{x_0}$, also called *mean squared prediction error* (*Cressie*, 1993), is given by:

$$\sigma^2(\underline{x_0}) = E\left[Z(x_0) - Z^*(x_0)\right]^2 = \sigma^2 - 2\sum_{i}^{N} \lambda_i Cov(h_{i0}) + \sum_{i}^{N}\sum_{j}^{N} \lambda_i \lambda_j Cov(h_{ij}) \tag{3}$$

According to the theory of *optimal linear estimators,* in order to evaluate λ_i, the estimator Z^* has to be *unbiased* (i.e. with the same expected value of Z) and such that to minimize eq. (3). These conditions lead to a system of $N+1$ equations:

$$\sum_{i=1}^{N} \lambda_i Cov(h_{ij}) + \mu = Cov(h_{0j}) \quad \text{j = 1, 2,…N} \tag{4.a}$$

$$\sum_{i=1}^{N} \lambda_i = 1 \tag{4.b}$$

where μ is a Lagrange multiplier that ensures the condition given by eq. (4.b). Once that the variance of the estimation error is known (eq. 3), the Root Mean Square Error in $\underline{x_0}$, can be computed as:

$$\Delta = \sqrt{\sigma^2(\underline{x_0})} \tag{5}$$

It is worth noticing that *kriging* method enables λ_i to be calculated on the basis of the geometry of the network. Indeed, once that the covariance is known, the distances between the stations and between the stations and the point where the estimation error needs to be computed, are sufficient to solve the *kriging* system (eqs. 4.a and 4.b).

2.3 Network selection

An adequate selection of the number of stations for a drought monitoring system can be carried out by computing the RMSE of the SPI index in points where no direct measurements are available. More specifically, the RMSE's corresponding to a given network configuration are computed in all points of a regular grid covering the whole region, yielding a map of the spatial distribution of the errors. Then, several network configurations can be analysed by repeating the procedure and by comparing the corresponding percentage of areas affected by given RMSE values.

For practical purposes it might be of some interest to assess, besides the RMSE, also the probability that the estimated SPI in a generic point is within a drought class (see table 1) given that the true value lies within a different class. In order to quantify such drought misclassification probabilities, the *method of fictitious point* (Delhomme 1978) can be applied to the proposed network. The method consists in repeatedly excluding a different station each time and assessing the value of the variable under study in that point based on the observations in the other *n-1* stations. By using this method, estimated series of SPI in each station are determined and compared with the observed series. Once that the number M of SPI values falling for both series in the same drought class is known, the probability of misclassification can be estimated as:

$$P = 1 - \frac{M}{L} \tag{6}$$

where L is the number of observations in the series. Although, the proposed methodology does not yield directly the optimal network configuration, however it enables, by a trial and error procedure, to identify a network configuration characterized by RMSE below a fixed limit. Further, the assessment of misclassification probabilities yields a measure of the network capability to correctly describe drought conditions over the investigated area. In what follows, first the procedure described in par. 2.2 is applied in order to identify a suitable network for drought monitoring; then a validation of such a network based on the probability of misclassification of the SPI index is carried out.

3. ANALYSIS OF SICILY HYDROMETEOROLOGICAL NETWORK

The proposed methodology has been applied to rainfall stations in Sicily. Since the automatic stations of the STIR network have been recently installed, no observed series are available yet. Therefore, in order to estimate the experimental covariogram, SPI series for each k have been computed based on rainfall data collected by 171 stations equipped with traditional instruments. In particular, besides being uniformly distributed all over the region under study, these stations are characterized by long historical series covering a period between 1921 and 1996 where occasional missing data have been filled by regression with neighbouring stations.

Once the SPI series for each station and time scale k have been estimated, the experimental covariograms can be computed and a theoretical model can be fit. Based on the shape of the experimental covariograms, the following theoretical model has been adopted:

$$Cov(h) = \sigma^2 \cdot \left[1 + \left(\frac{h}{b}\right)^2\right]^{-\beta}, \quad \beta > 0 \tag{7}$$

Figure 1 shows the experimental covariograms obtained, for different time scale k, by grouping couples of stations on the basis of distance intervals. In figure 1, for each plot two theoretical covariograms are represented: the first one interpolates the whole experimental covariograms, while the other one interpolates points corresponding to distances less than 100 km.

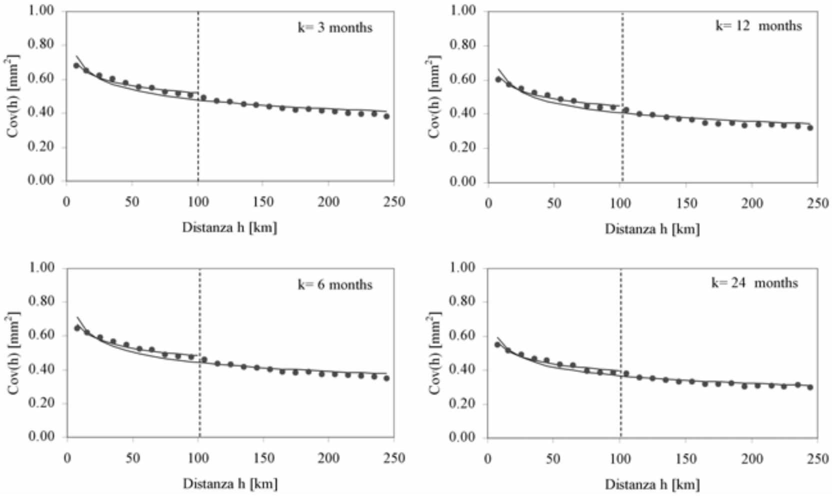

Figure 1. Covariance functions of SPI index

Since the RMSE is generally determined by using only the closest 4 stations, which usually lie within a distance of less than 100 km, it is worth referring to the latter covariogram. The parameters of such covariogram for each time scale *k* are:

Parameters	k=3 month	k=6 month	k=12 month	k=24 month
β	0.056	0.063	0.063	0.068
B	0.309	0.318	0.180	0.116

Then, a square grid, with cell size of 5 km, covering the whole study area, has been superimposed to the generic network configuration and estimation error in each node of the grid has been determined by means of eqs. (3) and (4). As an example, figure 2 shows a map of RMSE, obtained for SPI at *k*=24 months, corresponding to the configuration of 171 stations.

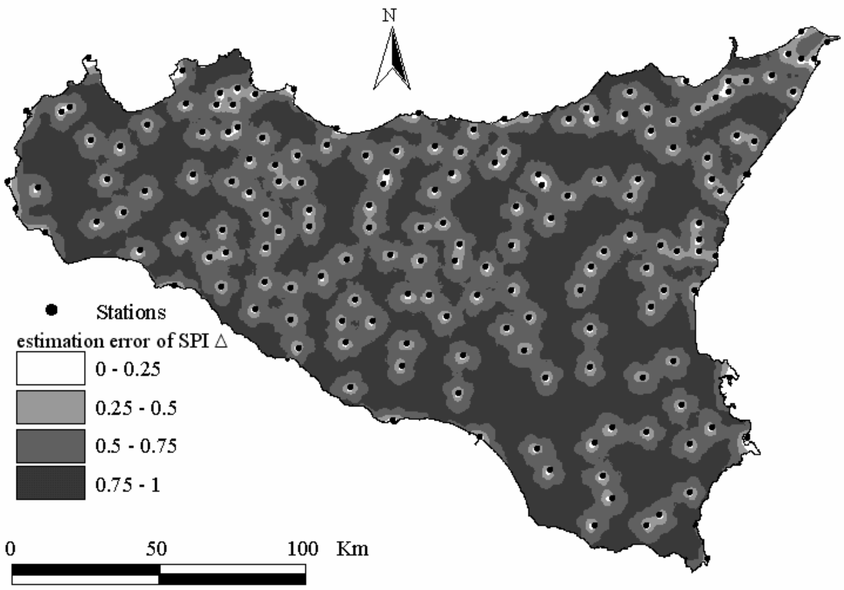

Figure 2. Map of the RMSE of SPI (k=24 months) for an hydrometeorological network with 171 rainfall stations

Among the 171 rainfall stations previously considered, only 102 belong to the automatic STIR network. Reminding that the objective of the investigation is to identify, within the automatic STIR network, a sub-network of stations specifically oriented to drought monitoring, the proposed methodology has been applied to such a configuration of 102 stations. Then, redundant gauges have been removed by a trial and error procedure, leading to a network of 90 stations.

In order to better summarize the results of the analysis, a comparison between areas (%) affected by different classes of error, either for the network of 171 stations or for the proposed sub-network composed by 90 stations, has been carried out for k=3,12 and 24 months.
Figure 3 shows some differences between the results of shorter periods *k* and those ones obtained for *k*=24 months. The class of RMSE which mainly affects the island, for *k*=3 and 12 months, is between 0.50 and 0.75 for both the considered configurations. By reducing the number of stations from 171 to 90, the largest class of error increases of about 10 % for *k*=3 months, while no significant changes occur for *k*=12 months. Instead considerable variations result for *k*=24 months for both configurations and an high increment of the areas related to the last two classes of error occurs by using the 90 stations network.

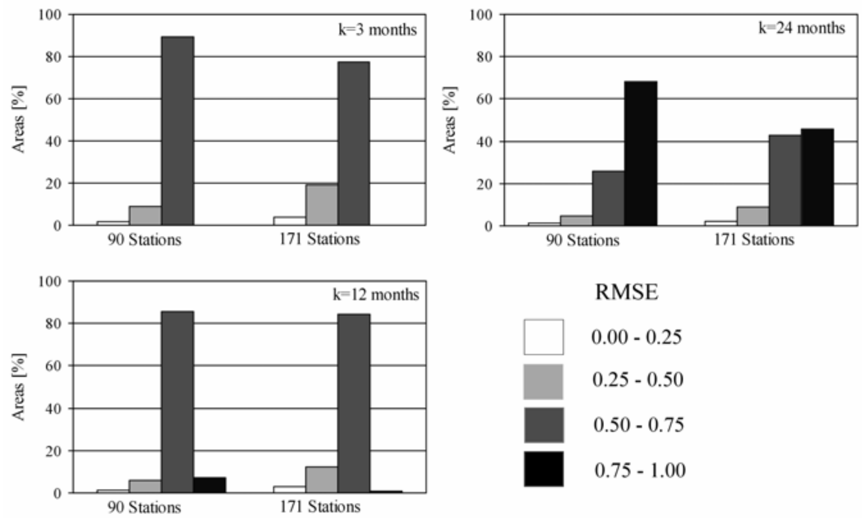

Figure 3. Comparison between areas within the same class of RMSE obtained for the two considered network configurations

The histograms representing the percentage of areas characterized by the different probability of misclassification, obtained by interpolating the values derived in the 90 stations with the method described above, are shown for *k*=3,12 and 24 months in figure 4. It can be observed that almost the whole region is affected by a probability of misclassification between 0.30 and 0.40. It is worth underlying that this probability includes all the possible misclassifications that can occur.

4. CONCLUSIONS

The increasing diffusion of automatic hydrometeorological networks requires a review of design criteria adopted for traditional networks, either in order to reduce installation and management costs of automatic gauges or to adapt networks to real time monitoring of extreme hydrometeorological phenomena. Furthermore, a uniform distribution of the stations over the region has to be coupled to the availability of long and reliable series of historical observations.
In this paper a methodology to select the adequate number of stations to be included in a network for drought monitoring has been presented. The proposed procedure is based on the determination by *kriging* of the RMSE of SPI index in points without measurement stations. In particular, a spatial distribution analysis of the RMSE obtained for various configurations has been carried out by determining the percentage of the areas affected by different classes of error. The application of the proposed methodology to the STIR network has enabled the identification of a sub-network for drought monitoring. Further a validation of such sub-network based on *the method of fictitious point* has shown that an estimated value of SPI in a generic point of the region might lead to an incorrect drought classification with a probability of 30÷40 % which although relatively high, is a direct consequence of the relevant morphological and climatic variability over the Sicily island.

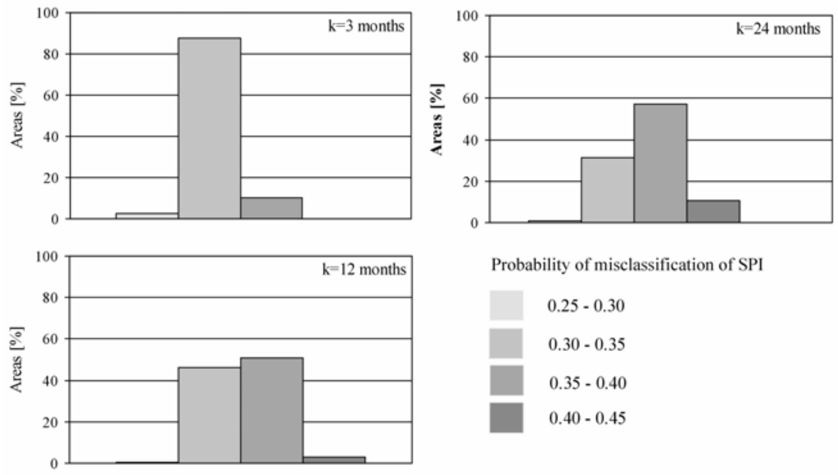

Figure 4. Distribution of areas characterized by different class of probability of misclassification of SPI obtained for the proposed configuration (90 stations)

It should be pointed out that the spatial distribution of the errors corresponding to the proposed configuration (90 stations, approximately 1 station per 260 km^2) is not significantly different than the one corresponding to the 171 stations, which represents the optimal network from a practical standpoint, considering the long data series availability. Further research is needed in order to take into account explicitly economic aspects related to the installation and management costs of the network.

ACKNOWLEDGEMENTS

This procedure has been developed within two researches founded by the European Project INTERREG IIc: the first one, "Reti di monitoraggio idrometeorologico per la previsione della siccità" (Contract n.116/1999), has been developed by DICA on behalf of Department of National Tecnical Service and the second one (Contract n.73/2001) on behalf of STIR-Sicily. Dr. Monacelli, Dr. Geraci and Dr. Vega are gratefully acknowledged for the useful discussions during all the stages of the investigation.

REFERENCES

Bacchi, B., 1996. Metodi di ristrutturazione e progetto delle reti pluviometriche. Gruppo Nazionale per la Difesa dalle Catastrofi Idrogeologiche, Presidenza del Consiglio dei Ministri, Dipartimento della Protezione Civile.

Bastin G., Lorent B., Dunqué C., Gevers M., 1984. Optimal estimation of the average areal rainfall and optimal selection of rain gauge locations. Water Resour. Res.; 20(4): 463-470

Bogardi I., Bardossy A., 1985. Multicriterion network design using geostatistics. Water Resources Research; 21(2): 199-208.

Cressie, N., 1993. Statistics for spatial data. Wiley & Sons, Inc. New York.

Delhomme J. P., 1978. "Kriging in the hydrosciences." In Adv. Water Resour.; l(5): 251-256.

Matheron G., 1965. Les variables régionalisées et leur estimation. M. et Cie (Ed) Paris.

McKee T., Doeskin N., Kleist J., 1993. The relationship of drought frequency and duration to time scales. Proceedings of the 8th Conference on Applied Climatology; January 17-22; Boston. American Meteorological Society.

Pardo-Iguzquiza E., 1998. Optimal selection of number and location of rainfall gauges for areal rainfall estimation using geostatistics and simulated annealing. Journal of Hydrology; 210: 206-220.