Measurement 113 (2018) 211–219

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

Retrospective analysis: A validation procedure for the redesign of an environmental monitoring network

E. Barca[a], D.E. Bruno[a], A. Lay-Ekuakille[b], S. Maggi[a,*], G. Passarella[a]

[a] CNR-IRSA, Water Research Institute, 70132 Bari, Italy
[b] Department of Innovation Engineering, University of Salento, 73100 Lecce, Italy

ARTICLE INFO

Keywords:
Groundwater monitoring
Monitoring network optimization
Kriging
Retrospective analysis

ABSTRACT

Monitoring networks are essential tools for the effective management of vulnerable or limited environmental resources. Cost and logistics constraints often suggest to reduce the number of monitoring sites while minimizing the loss of information determined by these changes. The problem can be rigorously addressed through the optimization of one or more objective functions that represent the managerial goals associated to the network. However, the use of objective functions is based on assumptions that in practical cases can be inaccurate. To overcome this problem, we have developed a retrospective analysis procedure that validates the degree of acceptability of the optimal reduced configuration at a local and global level. The procedure has been applied to a case study in Apulia, Italy, finding that the optimal reduced network was unable to recover the measured values of the monitored parameter of two discarded locations, making it unable to accomplish its monitoring goals.

1. Introduction

The optimal design of an environmental monitoring network is a key aspect of the effective and sustainable management of vulnerable or limited natural resources and much research effort has been spent to provide practical solutions to this problem.

An environmental monitoring network must be able to assess the state of a natural resource by reliably measuring a set of physical, chemical or biological parameters that characterize the system with the minimum amount of economic resources. In practice, these requirements correspond to maximizing the information content of the network while minimizing the costs and labor involved in the task.

For its inherent difficulty and practical usefulness, the problem has attracted the interest of several scientists, who have proposed a wide array of possible technical solutions [1–10].

The search of an optimal solution addresses two possible situations: the design of a new network and the redesign of an already existing network. The latter case is more frequent and also more challenging, since it can be very difficult to adapt an already existing network to new monitoring needs. Network redesign can either aim at increasing (*network upsizing*) or decreasing (*network downsizing*) the number of network monitoring sites over a given study area. A last and less common case consists in the rearrangement of the monitoring sites, while keeping unchanged the number of sites (*network relocation*), usually to increase the global efficiency of the network.

The redesign of a monitoring network can be performed by using one or more objective functions that capture the main features of the monitored parameters and the goals of the network management and assign a score to each of the possible configurations of the redesigned network. The network configuration that minimizes the objective function (or linear combinations of multiple objective functions) is considered the best possible redesigned network with respect to the selected features. Therefore, a great care must be taken in the proper choice of the objective function according to the managerial goals associated to the redesign of the network.

The spatial nature of monitoring networks and the successful application of geostatistics to spatial mathematical modeling problems, has paved the way to the use of objective functions based on geostatistical indices.

Kriging is a geostatistics interpolation technique that predicts the value of a parameter at locations where no measurement exists in terms of actual measurements at surrounding sites. A specific characteristic of kriging is that it associates to the prediction the kriging estimation variance (KEV), usually interpreted as a measure of the uncertainty of the prediction.

For this reason, objective functions based on KEV have been proposed for the optimization of monitoring networks [11–13]. The configuration that minimizes these objective function thus has the least

---

[*] Corresponding author.
E-mail addresses: emanuele.barca@ba.irsa.cnr.it (E. Barca), delia.bruno@ba.irsa.cnr.it (D.E. Bruno), aime.lay.ekuakille@unisalento.it (A. Lay-Ekuakille), sabino.maggi@cnr.it (S. Maggi), giuseppe.passarella@ba.irsa.cnr.it (G. Passarella).

http://dx.doi.org/10.1016/j.measurement.2017.08.014
Received 10 April 2017; Received in revised form 7 August 2017; Accepted 9 August 2017

overall estimation uncertainty among all the possible configurations of the redesigned network and the loss of information due to the network reduction can thus be considered negligible.

Unfortunately, experience with real-world datasets shows that the primary assumption of geostatistics, that similar values are always spatially clustered, might be too optimistic. Therefore, the use of objective functions based on geostatistical indices can be misleading. Since each monitoring site bears a precise piece of information, the effects produced by its removal cannot be always fully captured by the objective function.

In this paper, we propose to address this issue through a *retrospective analysis* validation procedure, that evaluates the capability of the reduced network configuration to recover most of the information of the original network. The analysis is performed at two different spatial scales: a first pointwise validation verifies the capability of the reduced network configuration to recover the measured values of the monitored parameters at each of the removed sites, and a global validation that evaluates the amount of measurement information lost by the reduced network over the whole study area. The effectiveness of the reduced configuration is quantitatively assessed by performing suitable statistical tests on the distributions of the set of measurements taken from the original network, and a second data set where the measurements at the removed sites are replaced by predictions based on the measured data at the remaining sites. If the network reduction is effective, the statistical tests should confirm the equivalence of the two distributions.

Based on the results of the validation stage, the monitoring network management can thus evaluate the balance between the loss of information and the cost savings provided by the network reduction and accept or reject, totally or partially, the new network configuration provided by the optimization process.

Retrospective analysis has been applied to a case study concerning the downsizing of the groundwater monitoring network composed of 61 piezometers, located in the shallow porous aquifer of the Tavoliere di Puglia, in southern Italy. The goal was to remove about 10% of the original sites with a minimal reduction of the information capability of the remaining network. Retrospective analysis found that, as desired, on a global scale the network reduction did not significantly affect the capabilities of the network. However, at a local scale the reduced network was unable to reliably recover the measured values of the monitored parameters at two of the discarded locations, suggesting the network management to reconsider the removal of these sites, since without them the reduced network could miss important local features of the monitored parameter.

2. Materials and methods

An environmental monitoring network is a physical system which measures one or more physical, chemical or biological parameters from sensors located at N sites, spread more or less uniformly over the area of the network.

A basic assumption behind the use of a monitoring network is that the measured parameters change continuously in space without abrupt changes at surrounding sites or, in other words, that the measured parameters are spatially auto-correlated [14]. If measuring a parameter at a monitoring site provides information about its value in its neighborhood, all sites falling within this area can to some extent be considered redundant.

Therefore, the spatial-autocorrelation property that makes it possible to set up a monitoring network, also makes it feasible to reduce the number of sites of the network without an appreciable loss of information.

For simplicity, in this paper we will limit the discussion to the optimal downsizing of a monitoring network that measures a single environmental parameter. The other cases of network redesign can be handled in a similar fashion.

2.1. Network downsizing

The downsizing process partitions the original set of N network monitoring sites into two separate subsets, the set of k removed sites and the set of remaining sites of the reduced network, composed of $N-k$ elements.

The number of possible sets of k removed sites is given by the binomial coefficient,

$$\binom{N}{k} = \binom{N}{N-k} = \frac{N!}{k!(N-k)!}, \quad 0 \leqslant k \leqslant N, \tag{1}$$

that diverges even for relatively small values of k and N. Therefore, for large N, the exhaustive exploration of all possible configurations to find the optimal reduced network (brute force search) is too demanding from a computational point of view to be feasible (except in the very special cases of $k \ll N$ or $k \approx N$ where the number of possible configurations becomes tractable) and clever algorithms to search the optimal solution must be used.

Classical optimization algorithms, such as gradient-based techniques, are much faster than brute force methods but can easily get trapped in local minima of the space of possible solutions, instead of finding the global minimum that represents the best solution to the problem.

Heuristic algorithms such as simulated annealing [15], genetic algorithms [16], particle swarm optimization [17] to name a few of the most widely-used algorithms, mimic physical or natural phenomena and represent a good trade-off between accuracy and computation time whenever classical methods are unable to find the globally optimal solution.

These algorithms can find the best solution within a predefined level of accuracy and in a reasonable computational time, without being trapped in a local minimum, by either applying a stochastic perturbation to the current state of the system and making it jump to different regions of the space of possible solutions (simulated annealing), or by exploring in parallel different regions of the solution space (genetic algorithms, particle swarm optimization).

2.2. Objective function

Let M be the set of size N, that represents the sites of the original monitoring network and k the number of sites to be removed from the network. The network downsizing process partitions M into two separate subsets, R and D, that represent the sites of the reduced network and the sites discarded from the network, respectively, where $M = R \cup D$.

Each set of monitoring sites is associated to the set of measured values of the monitored parameter z. Therefore, the set $M = \{x_1,x_2,...x_N\}$ of sites of the original network is associated to the complete set of measurements $Z = \{z(x_1),z(x_2),...z(x_N)\}$.

Similarly, $D = \{x_1^D,x_2^D,...x_k^D\}$, is associated to the removed measurements, $Z^D = \{z(x_1^D),z(x_2^D),...z(x_k^D)\}$ and R to a reduced set of measurements, $Z^R = \{z(x_1^R),z(x_2^R),...z(x_{N-k}^R)\}$, being $Z = Z^R \cup Z^D$.

The objective function for network downsizing problems can be expressed in the following general form,

$$\phi(\mathbf{D},\mathbf{Z^D}) = F(x_1^D,x_2^D,...x_k^D;z_1^D,z_2^D,...z_k^D). \tag{2}$$

The objective function $\phi(\mathbf{D},\mathbf{Z^D})$ can take different forms according to the managerial goals: for example it can try to optimize the spatial coverage of the network or the distribution of network sites with respect to the monitored area. Different objective functions will provide different reduced networks, since each of them is designed to accomplish a well-defined task.

A widely accepted classification splits them in *design-based* or *model-based* objective functions [18]. Design-based objective functions tend to be more accurate when the aim is to make some statistical inference about the spatial environmental parameter measured by the network,