ISBN: 978-92-990075-2-5

6th EnvImeko - IMEKO TC19
Symposium on Environmental Instrumentation and Measurements
June 24-25, 2016, Reggio Calabria, Italy

OPTIMAL RESIZING OF GROUNDWATER MONITORING NETWORKS: A RETROSPECTIVE ANALYSIS OF A REDUCED CONFIGURATION

E. Barca[1], D. E. Bruno[1], A. Lay-Ekuakille[2], S. Maggi[1], and G. Passarella[1]

[1] CNR-IRSA, Water Research Institute, 70132 Bari, Italy
[2] Dept. of Innovation Engineering, University of Salento, 73100 Lecce, Italy

Abstract: The redesign of environmental monitoring networks is a field rich of practical applications, in particular in light of recent EU environmental directives. When dealing with an existing monitoring network reduction, the easiest approach is to search for redundancies among the monitoring locations (clusters of locations) and to remove the most useless of them on the basis of the values of some quantitative parameter. This approach could lose some important features of the existing network, thus reducing its informative content. To avoid such a risk, the results of the network optimization should undergo a retrospective analysis capable to verify, by assessing some independent statistical indices, the acceptability of the reduced configuration with respect to specific managerial issues. The case study presented in this work focuses on the optimal downsizing of a groundwater monitoring network located in the Southern part of Italy. The MSANOS optimization software has been used to select the locations to remove from the existing network, with the goal to sustainably balance maintenance costs and information loss. The investigated case is of particular interest for both the critical issues of the considered resource and the recognized importance of the groundwater monitoring in the European Water Framework Directive.

Keywords: groundwater monitoring, monitoring network optimization, spatial simulated annealing, kriging variance estimation.

1. INTRODUCTION

The growing awareness of public opinion about issues related to environmental protection pushed the EU to rule about such important topic. In this framework, environmental monitoring assumes the fundamental role of a tool for assessing status and trends of natural resources and for defining and overseeing actions to recover impacted systems. Consequently, monitoring must be considered a strategic decision-making support tool and monitoring networks (MNs) should be designed as dynamic infrastructures, making them capable to adapt to any environmental change, for example by varying the size and the shape of the network.

In the last decades, the scientific and technical literature has proposed an extensive array of different methods for the optimal monitoring network redesign (OMNR) [1], using approaches characterized by different levels of complexity and detail. An effective monitoring network redesign should be based on a deep analysis of the characteristics of area, of the parameters to be monitored and of the available information (data-driven), with a clear idea of the monitoring objectives.

Given this basic knowledge, an OMNR problem consists of adding, removing or moving one or more monitoring sites within the area of interest. The problem can be solved by minimizing a suitable objective function (OF) which explores the network configuration space until specific conditions are met. The choice of the OF depends on the objectives and on the available information. Iterative optimization methods range from fully deterministic methods, for example jackknife or greedy deletion [2, 3], to stochastic methods, such as spatial simulating annealing [4]. Several software packages have been developed to solve this problem, using a wide array of different objective functions and optimization methods.

In this paper, the optimal downsizing of a long-term monitoring network is discussed. Each monitoring site bears a different piece of information and its removal could produce effects not always directly evaluable by the OF. Therefore, the impact of the removal of a site from the actual network on the capabilities of the MN should be investigated at different scales. We propose to address this issue with a retrospective analysis aimed to evaluate the capability of the residual network to recover the information lost at pointwise level and to estimate the population parameters at a regional scale. Some estimation error statistics are proposed in order to carry out this analysis.

The proposed method is applied to a case study concerning the downsizing of the groundwater level MN located in the shallow porous aquifer of the Tavoliere di Puglia, in southern Italy. Five monitoring stations have been chosen for removal from the original monitoring network composed of 61 piezometers. The downsizing of the MN has been carried out using the stochastic spatial simulating annealing (SSA) method, implemented into our MSANOS software package [5-8].

2. MATERIALS AND METHODS

A MN downsizing problem can be carried out by means of different combinations of methods and objective

ISBN: 978-92-990075-2-5

functions. The MSANOS software package used in this paper allows to set up a problem with regard to the specific available information and monitoring objectives. For the case under investigation, the ordinary average kriging variance (AKV) has been selected as the OF and the optimization problem has been solved by means of spatial simulated annealing (SSA).

The next subsections summarize the main characteristics of AKV and SSA.

2.1 Objective function

The AKV allows to assign a score to any network configuration in terms of average of the kriging variance estimated at any node of a discretization grid.

The well-known ordinary kriging variance formulation in a generic un-sampled location x_i is:

$$\sigma_R^2(x_i) = \sum_{j=1}^{N} \lambda_j(x_i)\,\gamma(x_j, x_i) - \mu(x_i) \qquad (1)$$

where $\gamma(x_j, x_i)$ is the variogram value for the (x_j, x_i) location pair, $\lambda_j(x_i)$ are the kriging estimation weights, $\mu(x_i)$ are the Lagrange multipliers and N is the size of the MN. Consequently, the objective function can be written as:

$$\phi_{AKV} = \frac{1}{w}\sum_{i=1}^{w} \sigma_R^2(x_i) \qquad (2)$$

where $w = r \cdot c$ is the number of grid points (rows × columns). The AKV is often chosen in spatial optimization problems because of its intrinsic capability of taking into account at the same time the local and the regional features of the variable of interest.

Obviously, as for any geostatistical based MN optimization, the variogram model is assumed to be known.

2.2 Spatial simulated annealing

The theory of SSA is inspired by the changes occurring in the atomic structure of a metal when it undergoes a process of quick heating followed by slow cooling (annealing), where the atoms of the metal change their arrangement and reach a configuration of lower energy. In this analogy, the spatial configuration of the atoms corresponds to that of the monitoring points, while the OF corresponds to the energy of the system.

Figure 1 summarizes the SSA algorithm, which basically consists of a pre-processing stage followed by two nested loops. In the pre-processing stage, the parameters that trigger the optimization method are estimated, namely the initial configuration S_0 and the initial temperature T_0.

The outer loop is controlled by the temperature, as in the physical analogy, and terminates when T approaches zero. The inner loop is related to the size of the problem: if k is defined as the number of locations to be removed, the inner loop consists of k iterations for a given value of temperature. Consequently, the maximum number of iterations of the method is $k\,N_T$, where N_T is the total of different values

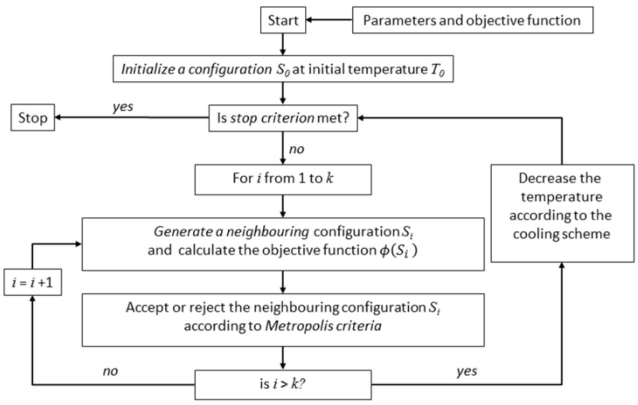

Figure 1. The SSA algorithm.

assumed by the temperature during the run. Within the inner loop, the candidate solutions are generated and subjected to the decision rule. For what is stated above, the total number of iterations corresponds to the total number of candidate network configurations. An important feature of the SSA is that it avoids falling in local optima.

2.3 Retrospective analysis of optimization results

Reducing the size of a groundwater MN is a decision-making process not free from risks. In general, whatever are the managerial objectives leading to the choice of reducing a MN, this task should remove the k less informative monitoring locations with respect to such objectives. At first glance, the OF should assure that the final network configuration is the best among all the possible candidates. Nevertheless, the correspondence between the managerial objectives and the OF is not always immediate and the actual impact of the sites removal on the residual informative capability of the MN should be further investigated. A retrospective analysis of the reduced configuration should then aim to a twofold purpose: (i) a local pointwise check to verify the reduced network capability of recovering removed observations (redundancy and spatial discontinuity); (ii) a global check to evaluate the overall impact of the removal of the locations on the spatial estimates at un-sampled locations over the considered area.

2.3.1 Local pointwise check

The relation between the original and the reduced monitoring network can be written as

$$M = M^{N-k} \cup D^k$$

where, M, M^{N-k} and D^k represent the original MN, the MN reduced by MSANOS by removing the k less informative sites and the set of the k removed sites, respectively.

This step of the retrospective analysis checks the capability of the reduced monitoring network to estimate the values observed at the removed sites. If successful, the test assures that the loss of information due to the MN reduction is negligible. The set of monitoring locations defined above can be associated to a corresponding set of observed values.

ISBN: 978-92-990075-2-5

Consequently, to the set $D^k = \{x_1^D, x_2^D, \cdots x_k^D\}$ of removed locations, corresponds a set of removed observations, $Z^D = \{z(x_1^D), z(x_2^D), \cdots z(x_k^D)\}$.

The main objective of the pointwise stage is to quantify the capability of the reduced network M^{N-k} and of its set of observations Z^{N-k} to recover the elements of Z^D estimated by kriging. The estimation of the removed observations is then carried out by means of the following equation,

$$\hat{z}(x_j^D) = \sum_{i=1}^{N-k} \omega_i\,(x_i)\, z(x_i), \;\; j = 1, 2, \cdots, k$$

where $\omega_i(x_i)$are the weights assigned to each $z(x_i) \in Z^{N-k}$. This test can be considered successful if the differences between the estimated and the observed values at the removed locations are negligible. To quantify the term negligible, one should consider the following empirical expression,

$$\left|\hat{z}(x_j^D) - z(x_i^D)\right| \leq \alpha \cdot z(x_i^D), \quad |\alpha| \leq 0.2$$

where α is a threshold of acceptability of the estimation error, usually set at 0.2. If, for each removed location, the difference between the estimated and observed values falls within the above interval, the monitoring site can be definitively considered redundant, and its information content can be reliably estimated by means of the observed values $z_i\,(x_i) \in Z^{N-k}$. In other words, if the estimations verify the test, it can be stated that the removed monitoring sites are redundant, since they can be successfully estimated from the remaining sites.

Another test that can be carried out at this stage of the retrospective local analysis, consists in comparing the observed and estimated values at the removed sites, on the basis of the confidence interval (CI) of the estimated values. The width of such CI can be easily derived as a multiple of kriging standard deviation [9]. If the observed values fall within the corresponding estimation CI at the given level of significance, they can be considered successfully estimated. This second local test is mainly devoted to check the presence of coarse local discontinuities at the removed sites, evaluating the coherence of the observed values at the removed sites with those at nearby locations. This occurs in practice when a small kriging standard deviation is associated to a large estimation error. In these cases, the optimization procedure should be run again, marking the considered site as not removable.

2.3.2 Global check

The removal of one or more sites may influence the capability of the monitoring network to describe the natural state of the environmental system taken as whole. A spatial estimation needs to be carried out over an estimation grid covering the whole monitoring network system.

Let $G^w = \{x_1, x_2, \cdots, x_w\}$ be the set of grid nodes, where $w = r \cdot c$ is the grid size. The parameter measured at the MN sites is computed twice at any node of the grid, using ordinary kriging. The first estimation is carried out considering the whole set of observations Z^N, while the second estimation uses only the reduced set Z^{N-k}. The two sets of estimations, $\hat{Z}^{G,N} = \{\hat{z}(x_1^{G,N}),\ \hat{z}(x_2^{G,N}),\ \cdots \hat{z}(x_w^{G,N})\}$ and $\hat{Z}^{G,N-k} = \{\hat{z}(x_1^{G,N-k}),\ \hat{z}(x_2^{G,N-k}),\ \cdots \hat{z}(x_w^{G,N-k})$, have the same size w.

Nevertheless, the homologous estimations will be different node by node, since the support to the estimations has changed. In any case, a reliable reduced MN should be capable of producing a set of estimations $\hat{Z}^{G,N-k}$ such that its empirical distribution is statistically equivalent to that of $\hat{Z}^{G,N}$.

Suitable statistical tests allow to infer the equivalence of the empirical distributions. Once that the normality and homoscedasticity (equal variance) of the distributions has been verified, the two-sample two-tailed t-test can be performed to compare the corresponding averages and to confirm the equivalence of the two distributions. In the other cases, the non-parametric Mann-Whitney test can be used to compare the averages.

In the former case, it is possible to confirm that the MN downsizing did not significantly affect the information capability of the network. In the latter case, even though the Mann-Whitney test is positive, the information capability of the reduced network results to be compromised, since the missing homoscedasticity indicates that it cannot reproduce the variability of the observed parameter with respect to the original MN.

3. CASE STUDY

The proposed case study refers to the downsizing of the groundwater level MN of the shallow porous aquifer of the Tavoliere di Puglia, located in the northern part of the Apulia region, Italy. The study area is a plain of about 3000 km^2, which is the largest alluvial plain in southern Italy (Fig. 2).

Three hydrogeological structures without vertical hydraulic connection can be identified here: a) a shallow porous aquifer in Quaternary alluvial and marine deposits, with depth less than 80-100 m; b) a porous aquifer in sandy portions within Pliocene clays, from about 100 m down to least 300 m below the surface; c) a deep karst aquifer in Cretaceous calcareous substratum, found at a depth of 300-600 m.

The shallow aquifer is located on alluvial gravel with interbedded sand and sandy-loamy sediments. The thickness of clay and silt layers increases in the external sectors of the alluvial basin, becoming aquitards that confine the aquifer layers. On the contrary, the coarse grained sediments prevail in the upstream sector of the aquifer located in the south-western part of the study area, in front of the Apennine range, where there is the main component of the aquifer recharge. The presence of silty-clayey sediments reduces the direct groundwater recharge, whereas an important recharge contribution comes as a base flow from the south-western unconfined sector of the aquifer, probably after a very long travel time (Fig. 3).

The Tavoliere di Puglia plain is mostly exploited for agricultural uses and the shallow porous aquifer represents the main source of water supply.

ISBN: 978-92-990075-2-5

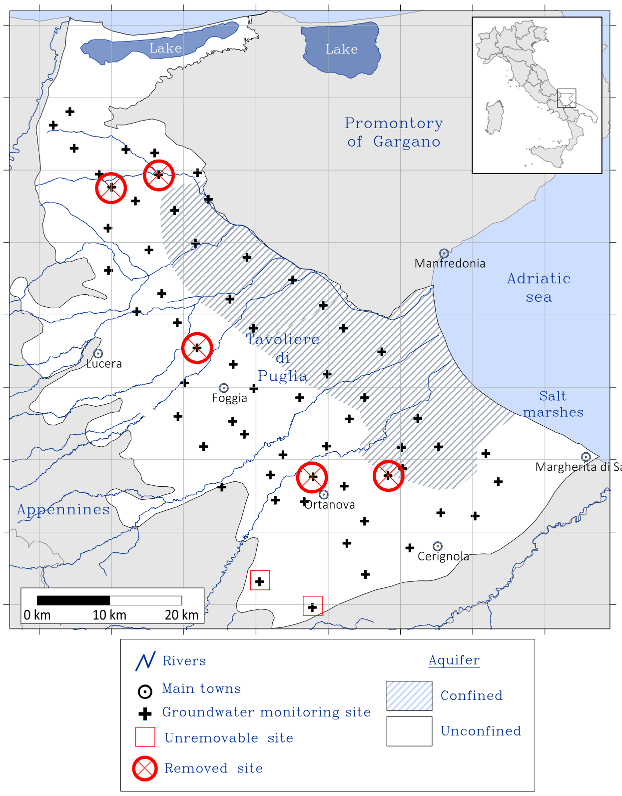

Figure 2. Study area.

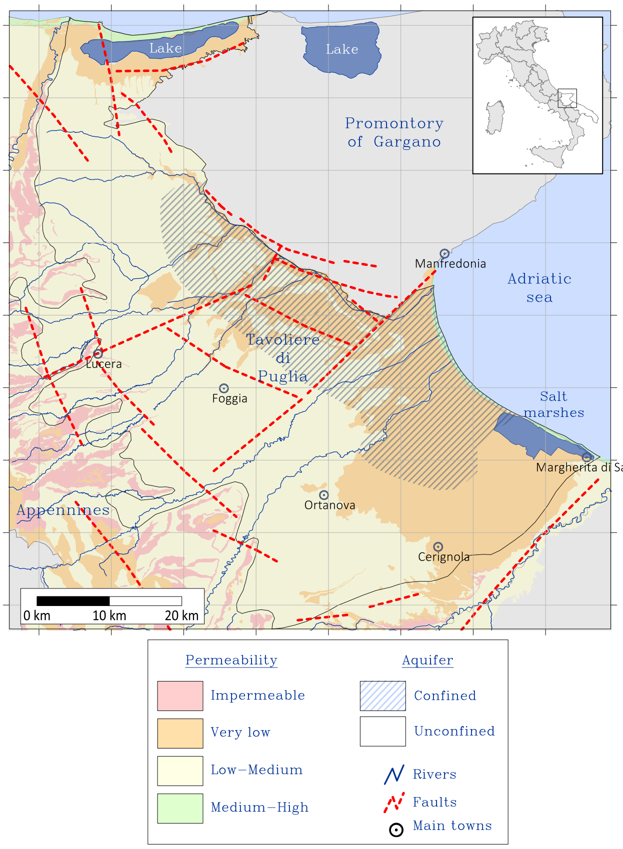

Figure 3. Tavoliere di Puglia: permeability map.

The aquifer size, hydrogeological-stress status, and hydro-geochemical characteristics make it an ideal experimental field in the framework of several national and international research projects [10-12]. Because of the intense exploitation of the Tavoliere di Puglia water resource, local authorities have been monitoring this water body since the 1990s. The MN consists of N = 61 wells, relatively well distributed over the area (Figure 2).

This network cannot be considered optimal since it was not specifically designed for monitoring purposes but was built by gathering already existing wells spread over the area. The main managerial objective is to regulate the access to the resource for agricultural uses. Nevertheless, to reduce the monitoring costs and to simplify the logistics of network maintenance, the local water authority aims to decrease the number of monitoring locations while minimizing the information loss and the increase of the uncertainty of the observed variable.

This is a typical constrained downsizing optimization problem that can be reliably addressed using an OF that is strictly related to the estimation uncertainty (AKV). In this framework, SSA was chosen because it guarantees the highest rate of success in converging towards the global optimum with respect to the other optimization methods.

The whole study area was covered with a 5 km × 5 km grid. The size of the grid was chosen to be sufficiently small with respect to the extent of the physical phenomenon of interest (variogram range).

Starting from the 61 measured values of groundwater level, the experimental variogram was calculated and a theoretical model was fitted after a thorough cross-validation stage. Table I reports the parameters of the variogram model.

The framework used was trans-Gaussian, since the variography was performed after a Gaussian anamorphosis of the distribution of the water table values In fact, a recent theoretical advance in geostatistics has proved the tight relationship between the linearity of the kriging interpolator and the Gaussian distribution of data [13].

In downsizing cases it is often useful, or even necessary, to keep some sites in the network. A specific feature of the MSANOS package allows to mark sites that should never be removed from the network. In this case, two wells located at the southern border of the study area were constrained as being not removable (Fig. 2). In fact, these two sites are located close to the boundary of the study area, where the kriging estimation variance is usually higher and in an area with a lower density of monitoring sites.

4. RESULTS AND DISCUSSION

Figure 2 shows the results of the optimization simulation carried out with the MSANOS software package. Cross-circled sites in the figure are those removed from the existing MN. The goal was to remove $k = 5$ wells,

Table I. Characteristic parameters of the variogram model.

Model	Nugget	Scale	Length
Gaussian	0.0350	1.1148	16000

ISBN: 978-92-990075-2-5

Table II. Pointwise estimation accuracy test.

Site	Observed	Estimated	Error	Relative error (%)
RW1	0.266	0.587	-0.321	120.8
RW2	-0.351	-0.179	-0.172	48.94
RW3	0.308	0.277	0.031	10.15
RW4	-0.625	-0.559	-0.066	10.50
RW5	-0.895	-0.748	-0.146	16.35

minimizing the information loss in terms of the AKV. The original network of 61 wells was characterized by an AKV value of 0.2941 m^2.

MSANOS provided the optimal reduced configuration in about 40 s after 2505 SSA iterations. The calculated AKV was 0.29671 m2. The efficiency loss is related to the difference between the original AKV_N and the final configuration's AKV_{N-k} . The relative percentage efficiency of the reduced network is

$$EI = \left(1 - \frac{AKV_{N-k} - AKV_N}{AKV_N}\right) \cdot 100 \qquad (3)$$

where N is the size of the original MN and k the number of removed sites. The resulting value of EI was 99.14%, meaning that, in terms of AKV, the loss of information is less than 1%.

Retrospective analysis gives an insight of the quality of the final reduced configuration. Local pointwise checks have been performed, followed by a global check of the overall impact of sites removal on the grid estimations.

The first local test concerns the pointwise estimation accuracy and consists in verifying if the estimation error is lower than the given acceptability threshold α, set here at 0.2 (20%). Table II lists the observed and estimated values of the variable of interest, together with the absolute and relative error for each of the removed sites. The last column of the table shows that the kriging estimations at two of the removed sites, RW1 and RW2, are affected by a very large error, much larger than the acceptability threshold. Therefore, this local check suggests to reconsider the removal from the network of RW1, and possibly also of RW2, since the dismissal of these sites could lead to missing local features of the variable of interest.

As a second local check, ordinary kriging has been used

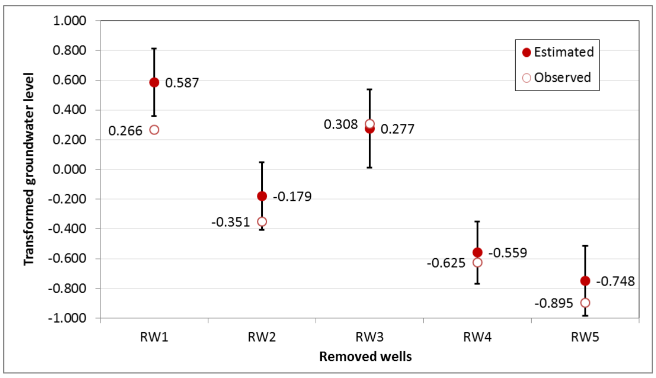

Figure 4. Comparison between the observed values and the values estimated from the reduced MN.

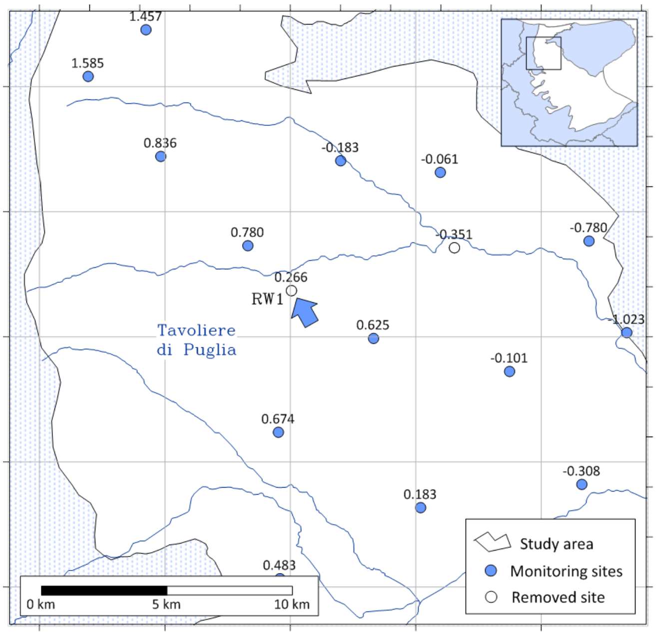

Figure 5. RW1 and its neighbouring sites.

to estimate the variable values and the corresponding prediction errors at the five removed sites. Afterwards, the CI of the estimated values, at the 96% level of significance (corresponding to twice the kriging prediction error), has been calculated, in order to compare these intervals to the observed values (Fig. 4). The figure clearly shows that four of the removed sites are certainly redundant, since their observed values are well within the CI of the estimations with the reduced network. The only exception is the first removed site, RW1, which is well outside the stated confidence interval. It is worth noting that, widening the CI at the 99% level of significance (three times the kriging prediction error), the estimated value in RW1 would satisfactorily match the corresponding observed value.

As already pointed out in Section 2, this test indicates that RW1 could be associated to a local spatial discontinuity, since its observed value is not coherent with nearby measurements. This interpretation is confirmed by Figure 5, that shows that RW1 is surrounded by several monitoring sites with significantly different measured values of groundwater level.

As a last global test, the consequences of the removal of the suggested sites on the residual informative capability of the whole MN has been assessed. Ordinary kriging was applied to estimate the variable of interest at each node of an estimation grid covering the whole study area, and not just at the removed locations, as in the previous case. The estimations were based on the observed value of both the original and the reduced MN.

Table III shows a summary statistic of the two resulting distributions of estimated values [14]. The values in Table III show a rather similar behavior of the two distributions, suggesting a negligible impact of the sites removal with regard to the global estimation of the considered variable. A more careful comparison between the two distributions has been also carried out by means of a t-test, confirming their substantial equivalence.

Table III. Summary statistics of the kriging estimations.

Statistics	Original network	Reduced network
Mean	0.018	0.020
Median	0.160	0.179
Standard deviation	0.928	0.933
Kurtosis	-0.280	-0.271
Skewness	-0.236	-0.215
Range	4.257	4.281
Minimum	-2.205	-2.199
Maximum	2.052	2.081
Mean standard error	0.086	0.086

5. CONCLUSIONS

A software package, MSANOS, was applied to a practical case of downsizing of an existing groundwater-level MN located in the Tavoliere di Puglia. Italy. SSA was used to optimize the selected AKV objective function. The aim of the study was to remove five wells, in order to reduce maintenance costs, while keeping most of the information capability of the network. MSANOS provided an optimal reduced MN, with about 99% of the original information capability. The quality of the optimization was checked by means of a retrospective analysis, which confirmed the negligible effect of the network downsizing on the global estimation capability. However, a local analysis marked at least one of the removed sites as critical, suggesting to reconsider such a removal. On the basis of these results, the MN manager decided to consider the RW1 site as un-removable The usefulness of the proposed retrospective analysis is clearly demonstrated by the case study.

ACKNOWLEDGEMENTS

The case study data has been collected during the TIZIANO Regional Groundwater Monitoring Project. The authors gratefully acknowledge the Water Protection Office of Regione Puglia for kindly allowing the use of the TIZIANO Project data.

REFERENCES

[1] E. Barca, A. Castrignanò, G. Buttafuoco, D. De Benedetto and G. Passarella, "Integration of Electromagnetic Induction Sensor Data in Soil Sampling Scheme Optimization Using Simulated Annealing," Environmental Monitoring and Assessment, vol. 187, no. 7, pp. 1-12, 2015.

[2] V. D'Agostino, G. Passarella and M. Vurro, "Assessment of the Optimal Sampling Arrangement Based on Cokriging Estimation Variance Reduction Approach," in Proc. of the XXV Congress of the International Association for Hydraulic Research, vol. 3, pp. 246-252, 1997.

[3] G. Passarella, M. Vurro, V. D'Agostino and M. J. Barcelona, "Cokriging Optimization of Monitoring Network Configuration Based on Fuzzy and Non-fuzzy Variogram Evaluation," Environmental Monitoring and Assessment, vol. 82, no. 1, pp. 1-21, 2003.

[4] A. Castrignanò, E. Barca, G. Buttafuoco, D. De Benedetto, D. A. Palumbo and G. Passarella, "Integration of Emi Sensor Data in Soil Sampling Scheme Optimization Using Continuous Simulated Annealing," in Proc. of GeoEnv 2014, 10th Conference on Geostatistics for Environmental Applications, N. Jeannèe and T. Romary, Eds., Presses des Mines, Collection Sciences de la Terre, Paris, France, pp. 34-35, 2014.

[5] E. Barca, G. Passarella and A. Morea, "A Software for Optimal Information Based Downsizing/Upsizing of Existing Monitoring Networks," in Proc. of Spatial Data Methods for Environmental and Ecological Processes, 2nd Edition, B. Cafarelli, Ed., 2011.

[6] E. Barca, G. Passarella, M. Vurro and A. Morea, "Optimal Redesign of Groundwater Monitoring Networks: a Data-driven Multi-approach Software," in Proc. 8th International Conference of EWRA, Water Resources Management in an Interdisciplinary and Changing Context, Porto, Portugal, 2013.

[7] E. Barca, M. C. Calzolari, G. Passarella and F. Ungaro, "Predicting Shallow Water Table Depth at Regional Scale: Optimizing Monitoring Network in Space and Time," Water Resources Management, vol. 27, no. 15, pp. 5171-5190, 2013.

[8] E. Barca, G. Passarella, M. Vurro and A. Morea, "MSANOS: Data-driven, Multi-approach Software for Optimal Redesign of Environmental Monitoring Networks," Water Resources Management, vol. 29, no. 2, pp. 619-644, 2015.

[9] J. P. Chilès and P. Delfiner, Geostatistics: Modeling Spatial Uncertainty, Wiley, New York, 1999.

[10] E. Barca, G. Passarella, R. Lo Presti, R. Masciale and M. Vurro, HarmoniRiB River Basin Data Documentation, Ch. 7, HarmoniRiB Project Deliverables D6.3, pp. 1-47, 2006.

[11] E. Barca, G. Passarella, R. Lo Presti, R. Masciale and M. Vurro, Candelaro River Basin, Italy: A HarmoniRiB case study, HarmoniRiB Project Deliverables D7.6, pp. 1-54, 2006.

[12] R. Masciale, E. Barca and G. Passarella, "A Methodology for Rapid Assessment of the Environmental Status of the Shallow Aquifer of 'Tavoliere di Puglia' (Southern Italy)," Environmental Monitoring and Assessment, vol. 177, no. 1-4, pp. 245-261, 2010.

[13] P. Diggle and P. J. Ribeiro, Model-Based Geostatistics, Springer Series in Statistics, Springer, Berlin, 2007.

[14] E. Barca, E. Bruno, D. E. Bruno and G. Passarella, "GTest: A Software Tool for Graphical Assessment of Empirical Distributions' Gaussianity," Environmental Monitoring and Assessment, vol. 188, no. 3, pp. 1-12, 2016.