ISBN: 978-92-990075-2-5

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

HEURISTIC RULES FOR A RELIABLE VARIOGRAM PARAMETER TUNING

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, Via F. De Blasio 5*

[2]*Department of Innovation Engineering, Università del Salento, 73100 Lecce, Italy, Via Monteroni*

Abstract: The variogram incorporates the basic information for the geostatistical analysis of spatial data, making it one of the most popular and widespread approaches for the evaluation and assessment of environmental data. Fitting an experimental variogram with a proper model, capable of catching the main characteristics of the sampled data, is a critical stage of spatial analysis. This stage is generally carried out using a semi-automatic approach, with different levels of user's contribution. Once a model and the related starting parameter values have been defined, they can be refined using a trial-and-error strategy and the impact of such changes can be evaluated through several calibration indices. In this paper, we present an overview of the main indices used for the refinement of the variogram model parameters, highlighting the specific information provided by each of them together with a few heuristic strategies for driving the trial-and-error refinement process.

Keywords: geostatistics, structural analysis, variogram model parameters, heuristic rules.

1. INTRODUCTION

Data sampled at specific spatial locations usually show an evident spatial correlation, as stated by the well-known Tobler's first law of geography, *everything is related to everything else, but near things are more related than distant things* [1]. Taking advantage of the aforementioned law, it is possible to expand a small set of sampled values and to estimate the values of the variable of interest at nearby locations from the observed spatial data, with the ultimate goal of up-scaling the knowledge of these properties from single spatial locations to larger areas [2]. Geostatistics is the main methodological framework that provides suitable methods for modeling the spatial behavior of a specific variable.

The main tool for capturing the features of a spatial property is the variogram model, a functional law that describes the structural and random components of a spatially autocorrelated variable. The construction of a variogram model is commonly achieved through three stages: computation of the experimental (or empirical) variogram, selection of an acceptable theoretical model, fitting of the theoretical model to the empirical variogram. A theoretical variogram model is a *conditionally negative definite* function [3, 4]. Since the empirical variogram does not usually satisfy such a property, it is not straightforward to derive from it the theoretical model, complicating the task of capturing the underlying spatial autocorrelation from the sampled data [5].

Such a task is commonly accomplished using one of the following strategies: (i) trial-and-error, where the parameters of the selected theoretical model are changed manually, checking the effect of each change on suitable calibration indices; (ii) Bayesian methods, which allow a limited user intervention during the optimization stage; and lastly (iii) automatic optimization methods, such as ordinary least squares (OLS), weighted least squares (WLS), maximum likelihood (ML) and restricted maximum likelihood (REML) methods [6]. Bayesian or automatic optimization methods are usually used to define a possible theoretical model and the related initial parameter values. The model is then further refined by trial-and-error, and the effect of the changes to the model parameters is evaluated through several statistical calibration indices (accuracy indices). The trial-and-error strategy uses the (Leave-One-Out) Cross-Validation procedure for computing the calibration indices and tuning the variogram model parameters. Such procedure iteratively omits one of the sampled data and estimates the variogram from the reduced set of observations, computing indices based on the residuals. If the variogram model is optimal, the indices meet some predefined numerical criteria, otherwise the variogram model parameters are changed until the results satisfy such criteria. If the expected criteria cannot be reasonably met by adjusting the parameters of the theoretical variogram function, calibration needs to be reinitialized using a different theoretical model.

The paper focuses on the refinement of the theoretical variogram model using the trial-and-error strategy. We introduce the main calibration indices used for the refinement of the model parameters and describe the heuristic strategies for manually improving the agreement of the model to the experimental data. The application of the trial-and-error strategy to a specific case study is also briefly reported.

2. TRIAL-AND-ERROR STRATEGY

2.1. Variogram Model Parameters

The main parameters of a variogram model are the *nugget*, the *length* and the *scale* (Fig. 1). The nugget accounts for

ISBN: 978-92-990075-2-5

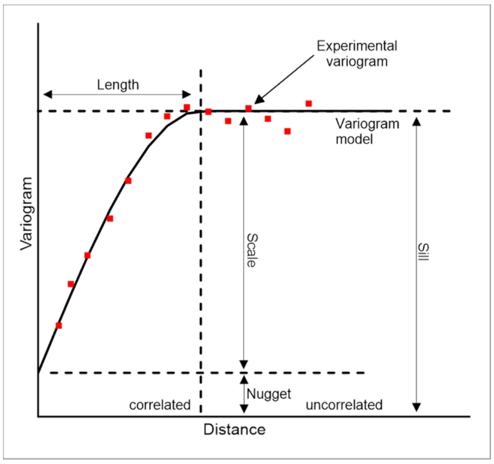

Figure 1: Main parameters of a variogram model.

the data uncertainty at small scale, the length represents the distance beyond which the auto-correlation vanishes and the scale indicates the variability assigned to the model, which should be approximately equal to the variance of the experimental data. Other useful model parameters are the *range*, which is proportional to the length, and the *sill*, defined as the sum of the nugget and the scale. These parameters are refined by trial-and-error in order to improve the fit of the variogram model to the observed data.

2.2. (Leave-One-Out) Cross-Validation Procedure

Let Z be the set of N experimental observations, $\{z(x_i),\ i=1,\dots N\}$. From this set, N reduced sets of size $N-1$ can be generated. For $i=1$ up to N, the i-th reduced set Z_{-i} is obtained by dropping the i-th observed value $z(x_i)$ from the full set,

$$Z_{-i} = \{z(x_1), z(x_2), \cdots z(x_{i-1}), z(x_{i+1}), \dots z(x_N)\}. \quad (1)$$

The estimation of the i-th observed value is performed using the *kriging method*, as follows

$$\hat{z}(x_i) = \sum_{j=1}^{N-1} \lambda_j(x_j) z(x_j), \quad (2)$$

where $\lambda_j(\cdot)$ is the generic kriging weight and $N-1$ is the size of the reduced set of observations. The kriging variance associated to the estimated value is

$$\sigma_K^2(x_i) = \sum_{j=1}^{N-1} \lambda_j(x_j)\gamma(x_j, x_i) - \mu(x_i), \quad (3)$$

where $\lambda_j(\cdot)$ is the generic kriging weight, $\gamma(\cdot)$ is the variogram value associated to the pair (x_j, x_i) and $\mu(x_i)$ is the Lagrange multiplier [4]. At the end of the cross-validation procedure, a number of statistical calibration indices are computed.

2.3. Calibration Indices

The cross-validation kriging error (or residual), r_i, is defined as

$$r_i = z(x_i) - \hat{z}(x_i), \qquad i = 1, \dots N, \quad (4)$$

where $z(x_i)$ are the observations and $\hat{z}(x_i)$ are the estimations.

A first criterion for assessing the goodness of the cross-validation results is the *unbiasedness condition*, that requires that the mean-bias error (MBE) is close to zero [7],

$$\text{MBE} = \frac{1}{n}\sum_{i=1}^{n} r_i \approx 0, \quad (5)$$

balancing under-estimations and over-estimations during the cross-validation procedure.

Probabilistic statements concerning the actual value of z at an unmeasured location and its kriging estimate and kriging variance at that location, can be made by analyzing the distribution of the cross-validation kriging residuals.

Therefore, an analysis of the distribution of the residuals should be done to check for the Gaussianity with zero mean of such a distribution.

A second reasonable criterion for selecting a theoretical variogram is to minimize the sum of the squared errors,

$$\sum_{i=1}^{n} r_i^2, \quad (6)$$

with respect to the variogram parameters. Unlike ordinary least squares regression, the minimization of the sum of the squared errors (6) is not sufficient for kriging because the resulting model can yield inconsistent estimates of the kriging variance, $\sigma_K(x_i)$, at location x_i, giving unrealistic measures of the accuracy of the kriging estimates. To prevent such a bias, one should use the *reduced kriging residual* [4],

$$r_i' = \frac{r_i}{\sigma_K(x_i)}. \quad (7)$$

If the kriging variance is a consistent estimate of the true mean-squared error (MSE) of estimate, then the standard deviation of the reduced kriging residuals should be close to 1.0.

Therefore, the standard cross-validation technique for evaluating a theoretical variogram requires the minimization of the root-mean-squared error (RMSE) [8, 9],

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} r_i^2}, \quad (8)$$

subject to the condition that the reduced root-mean-squared error (RRMSE)

$$\text{RRMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[\frac{r_i}{\sigma_K(x_i)}\right]^2} \approx 1. \quad (9)$$

The requirement in Eq. (9) can usually be relaxed, requiring that the RRMSE falls within the interval $1-2\sqrt{2/n}$ and $1+2\sqrt{2/n}$ [10]. It is also important to check that the distribution of the reduced residuals is Gaussian, with zero mean and variance equal to one.

The average standard error (ASE) is defined as

$$\mathrm{ASE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[\sigma_K(x_i)]^2}. \quad (10)$$

A numerical condition almost equivalent to the condition imposed on RRMSE, Eq. (9), can be imposed on the ratio between ASE and RMSE,

$$\frac{\mathrm{ASE}}{\mathrm{RMSE}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[\sigma_K(x_i)]^2} \bigg/ \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[\frac{r_i}{\sigma_K(x_i)}\right]^2} \approx 1. \quad (11)$$

It is well known that MBE suffers of the *cancellation effect*, due to the reciprocal cancellation of positive and negative residuals. Therefore, MBE cannot be considered a valid estimator of the average residual magnitude. A reliable estimation of the residual magnitude should be based on different indices, such as RMSE, Eq. (8), or the mean average error (MAE), defined as

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|r_i|. \quad (12)$$

In principle, the values of RMSE and MAE should be equal. However, as pointed out by several authors, the squaring effect in RMSE magnifies the larger residuals, resulting in RMSE $\geq$ MAE. As a rule of thumb, a ratio MAE/RMSE $\geq$ 0.8 indicates a good performance of the selected variogram model [9].

A last index accounting for the relative magnitude of the average of the error (RMAE) is needed in order to consider the magnitude of the error in relation to the observed value,

$$\mathrm{RMAE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|z(x_i)-\hat{z}(x_i)|}{z(x_i)} = \frac{1}{n}\sum_{i=1}^{n}\frac{|r_i|}{z(x_i)} \quad (13)$$

An excellent variogram model gives an RMAE of about 10%, but a value of $15-20\%$ can still be considered acceptable.

2.4. Tuning of Model Parameters

The length is the main parameter to be tuned in order to refine the theoretical variogram model. The tuning of the length parameter affects the magnitude and sign of the kriging weights and also changes the kriging variance.

In general, the impact of the length on the magnitude of the kriging weights is small. However, even small changes in the kriging weights can cause significant effects on kriging estimations. Since there are no explicit constraints on the sign of the kriging weights, they can potentially become negative at high values of the length, or when the search ellipse that defines the neighborhood of the site being estimated becomes too wide. Negative weights can distort the estimate, but this issue can be easily addressed by decreasing the value of the length.

Lastly, the increase of the length decreases the ordinary kriging variance, since it increases the neighborhood of the estimated site. Conversely, when the length becomes very small, all the known values fall outside the neighborhood of the estimated point. This leads to the undesirable case of a kriging estimation that becomes a simple average, with weights equal to $1/(N-1)$.

The degree of spatial auto-correlation is usually inversely related to the nugget-to-sill ratio. Therefore, the size of the nugget should be always thought as a percentage of the sill and not as an absolute value. Decreasing the nugget affects the kriging variance and prevents the introduction of a systematic error in the kriging estimation. Conversely, increasing the value of nugget leads to a more even distribution of the kriging weights and to an increase of the kriging variance. If the value of the nugget approaches that of the sill, there is no information redundancy between the two values. The result is a simple average of the available data with a complete lack of spatial correlation. Changing the value of the sill does not affect the ordinary kriging weights and the estimated value (if the nugget is $= 0$), but affects the kriging variance and the resulting error estimation. A change of the sill changes by the same factor also the kriging variance.

3. VARIOGRAM MODEL REFINEMENT

The tuning of the variogram parameters is quite holistic and involves the analysis of the calibration statistics as well as of several graphical plots, in order to widen the insight of the structure underlying the data.

The main graphical tool is the variogram plot, which allows to visually check the degree of agreement between the model and the experimental variogram. The variogram model should fit as much as possible to the experimental variogram, capturing its main spatial features and at the same time avoiding over-fitting, which would add to the model inessential and erratic characteristics of the experimental variogram. The change of any of the model parameters has a strong effect on the variogram plot. Therefore each change must be visually inspected in order to correct excessive departures of the model from the experimental plot.

The Q-Q plot of the reduced residuals compares the normal quantiles with those corresponding to the reduced residuals distribution. If the points on a Q-Q plot are arranged along a straight line in the first quadrant bisector, the distribution of the reduced residuals can be considered Gaussian, with zero mean. Marked deviations from linearity provide evidence against this hypothesis.

Another useful graphical tool is the variogram plot of the residuals. When this plot does not show a spatial structure of the residuals, it means that the variogram model of the data has captured the underlying structure in its entirety.

The last tool considered here is the post-plot, which shows the position of the residuals together with their values. This plot is particularly useful because it highlights the effect of tuning the variogram parameters on the estimations.

Before dealing with the heuristics of parameter tuning, it is worth noting that the goal of parameter adjustment is to improve cross-validation statistics and should be carried out only if the indices depart unacceptably from their limit values. All modifications to the parameters should avoid to lead to a drastic deviation of the theoretical model from the experimental variogram. If the support of the sampled points cannot achieve acceptable cross-validation results within the given variogram model, a different model should be investigated.

The calibration indices described above can be divided into two separate groups: *sensitive* and *inertial* indices. The first group is sensitive to small changes of the model parameters, and comprises MBE, RRMSE and ASE. The inertial statistics indices, RMSE, MAE and RMAE, are usually less affected by modifications to the model parameters.

When sensitive calibration indices assume unacceptable values, they can usually be improved by varying the length or the nugget. Small changes to the length parameter have a direct impact on the MBE and the RRMSE. If RRMSE is too small, $\ll 1$, increasing the length (equivalent to decreasing the slope near the origin of the variogram) decreases the kriging variance and increases the value of RRMSE. The converse is true when the RRMSE is too large, $\gg 1$.

If changes to the length do not significantly improve the RRMSE calibration index, the tuning of the nugget parameter should be considered. Geometrically, the nugget is the intercept of the variogram model with the $y-$axis. The nugget represents the variance near the axis origin and is always non-negative. A decrease of the nugget decreases the kriging variance and, consequently, the value of the average standard error ASE. When the MBE is negative, an increase of the nugget can adjust this statistics, while for positive values of MBE, a decrease of the nugget should be considered.

As already noted, the value of the scale parameter should be approximately equal to the data variance. In general, the increase of this parameter increases the ASE and decreases the RRMSE.

Considering the inertial statistics indices, when MBE > 0, the only way to decrease the magnitude of RMSE and MAE and the relative magnitude RMAE is to reduce the value of the nugget. In fact, the larger the ratio between the nugget and the sill, the larger the values of the inertial statistics indices. When MBE < 0, an increase of the nugget can improve such index. It is worth considering that improvements to one index usually occur at the expense of the other indices, therefore that a particular care should be taken to properly balance the various statistical indices, for example by following a law of *conservation of variance*, where on average the variance of the data should be reproduced by the kriging variance. Therefore, to avoid both over-estimation and under-estimation of the kriging variance, the value of RMSE should be similar to that of ASE.

Gaussian distribution checks should always be carried out for the residuals and the reduced residuals. In particular, the reduced kriging errors should approximate a standard normal distribution. If this is not true, a transformation of the data might be needed to achieve a more normal distribution, followed by a new variogram estimation procedure.

The Chauvenet criterion can be useful to identify outliers or badly located points (i.e., not robust points), which can negatively affect the variogram model parameters. Such points should be discarded (or at least adjusted) from the data set [11]. Changes to the experimental variogram due to the removal of the outliers also require to repeat the whole variogram estimation procedure.

variogram parameter	automatic model	refined model	units
length	14939.48	16000.00	m
range	25875.94	27712.81	m
nugget	0.047	0.035	m^2
scale	1.048	1.115	m^2
sill	1.095	1.150	m^2

Table I: Variogram model parameters calculated (a) with an automatic OLS optimization method and (b) after refinement with the trial-and-error strategy.

4. APPLICATION TO A CASE STUDY

The trial-and-error strategy has been assessed using a data set consisting of piezometric head measurements [12, 13]. A Gaussian variogram model has been calculated using an automatic OLS method, obtaining a set of initial model parameters and the corresponding statistical indices. A thorough structural analysis has preliminary been carried out to select the proper variogram model among several possible isotropic and anisotropic alternatives. The best agreement with the measured data has been obtained with the selected isotropic model.

A trial-and-error strategy has been applied to improve the values of the calibration statistics, obtaining a new set of refined model parameters.

The results are summarized in Table I and Table II, that list the model parameters and the corresponding statistical indices before and after refinement. The plots of the initial and refined variogram models are shown in Figure 2, while Figure 3 shows the Q-Q plots and the post-plots of the initial and refined models.

The automatic model fitting strategy performs quite well with this data set, as it is evident by the values of the statistical indices of Table II, in particular MBE, RMSE and MAE.

statistical index	automatic model	refined model	% relative improvement
MBE	−0.00150	0.00036	124.0
RMSE	0.25906	0.25581	1.3
ASE	0.29927	0.25767	13.9
MAE	0.19890	0.19652	1.2
RMAE	0.77509	0.76982	0.7
RRMSE	0.88233	1.02203	118.7
ASE/RMSE	1.15522	1.00727	95.3
MAE/RMSE	0.76778	0.76823	0.2

Table II: Calibration indices of (a) the initial variogram model calculated with an automatic OLS optimization method, and (b) of the model refined with the trial-and-error strategy. The last column shows the relative improvement of the refined statistical indices with respect to the limit values of each index.

ISBN: 978-92-990075-2-5

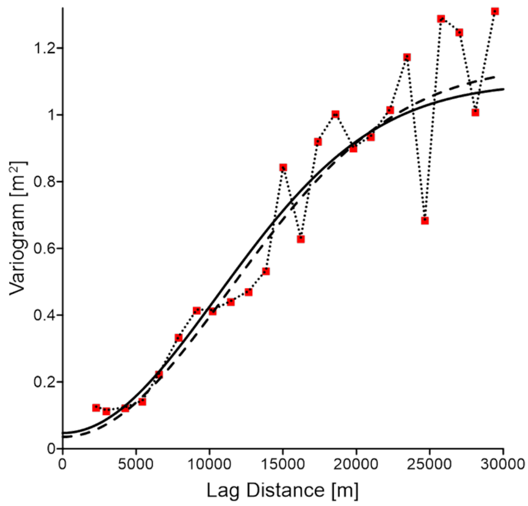

Figure 2: Experimental variogram (red squares) together with the initial (solid line) and refined (dashed line) variogram models calculated from the measured data set.

However, the values of the ASE/RMSE and MAE/RMSE ratios are not so close to their limit values (Table II), suggesting the possibility of a further improvement of the variogram model parameters.

Increasing the length and decreasing the nugget by small amounts dramatically improves some statistical indices, in particular MBE, RRMSE and ASE/RMSE and, to a lesser extent, ASE, as shown in the refined column of Table II. Other indices, namely RMAE and MAE/RMSE, change very slightly and are almost insensitive to modifications of the model parameters. Also the Q-Q plot of the refined model shows only a minor improvement, since the reduced residuals of the initial model are already very well aligned along the reference line (Fig. 3). Inspecting the post-plot of the refined model, it is apparent that the size of the posts, that is proportional to the absolute magnitude of the residuals, has been reduced by the refinement (Fig. 3), proving the effectiveness of the trial-and-error strategy for tuning the variogram model parameters.

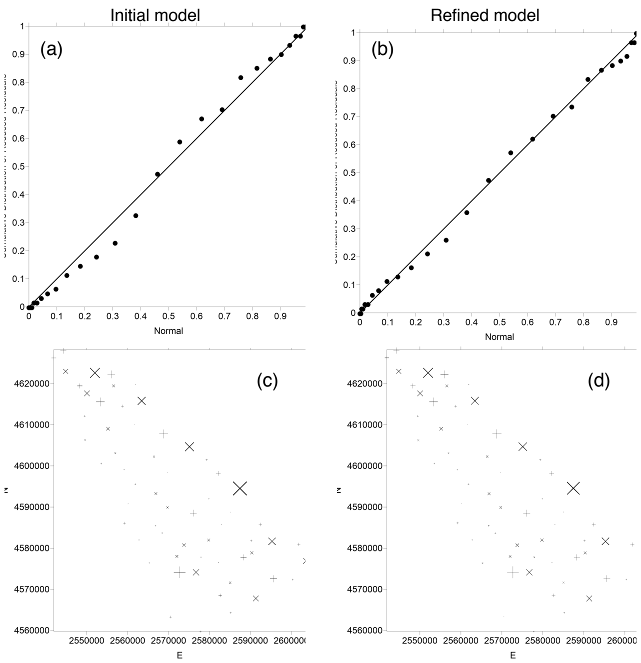

Figure 3: Graphical plots of the initial and refined variogram models calculated from the measured data set: (a)-(b) Q-Q plots, (c)-(d) post-plots.

5. CONCLUSION

The final and most important stage in the construction of a theoretical variogram model is the refinement of the model parameters, in order to improve the agreement of the theoretical model to the available experimental data. We have reviewed the main statistical indices used to evaluate the effect of the changes to the model parameters, highlighting the specific information provided by each index. We have also described a few heuristic strategies for manually improving by trial-and-error the agreement of the model to the data. The application of the trial-and-error refinement strategy to a case study is also briefly reported.

ACKNOWLEDGMENT

The data used in the case study 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] W. Tobler, “A computer movie simulating urban growth in the Detroit region”, Economic Geography, vol. 46, no. 2, pp. 234–240, 1970.

[2] A. G. Journel and C. J. Huijbrechts, Mining Geostatistics, Academic Press, 1978.

[3] J. P. Chiles and P. Delfiner, Geostatistics, Modeling Spatial Uncertainty, Wiley, 1999.

[4] E. H. Isaaks and R. M. Srivastava, An Introduction to Applied Geostatistics, Oxford University Press, 1989.

[5] 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.

[6] N. A. C. Cressie, Statistics for Spatial Data, Revised Edition, Wiley, 1993.

[7] T. Chai and R. R. Draxler, “Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature”, Geoscientific Model Development, vol. 7, no. 3, pp. 1247–1250, 2014.

ISBN: 978-92-990075-2-5

[8] C. J. Willmott and K. Matsuura, “Advantages of the Mean Absolute Error (MAE) Over the Root Mean Square Error (RMSE) in Assessing Average Model Performance”, Climate Research, vol. 30, pp. 79–82, 2005.

[9] C. J. Willmott, K. Matsuura and S. M. Robeson, “Ambiguities Inherent in Sums-of-Squares-based Error Statistics”, Atmospheric Environment vol. 43, pp. 749–752, 2009.

[10] J. P. Delhomme, “Kriging in the hydrosciences”, Advances in Water Resources, vol. 1, no. 5, pp. 251–266, 1978.

[11] W. Chauvenet, A Manual of Spherical and Practical Astronomy Vol. II (Lippincott, Philadelphia, 1863), Reprint of 1891 5th Edition, Dover N.Y., pp. 474–566, 1960.

[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, 2011.

[13] 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.