Univercity of Crete
Department of Mathematics and Applied Mathematics

Groundwater monitoring networks management using Genetic Algorithms

by
Antonios E. Parasyris

Under the supervision of Dr. N. Kampanis

A thesis submitted in partial fulfillment for the Master's Degree
September-June 2016

This page intentionally left blank

Contents

List of Figures

List of Tables

Acknowledgments

Firstly, i would like to thank Dr. N. Kampanis for his supervision, his mentorship, and because without him, i would not be part of the Coastal Research Laboratory, where i met wonderful people, to whom i owe special thanks. More specifically, for the continuous help to this work, her constant advice and her idea/motivation for this work, special thanks goes to Dr. K. Spanoudaki. Also,for his insights in Geostatistics and his willingness to help in every chance, i would like to thank Dr. E. Varouchakis. Without Katerina and Manolis crucial help, this work would be completely different. For their guidance and for the doors that i found wide open for every question, i would like to thank each and every one of my professors at the department of Mathematics in University of Crete, in both undergraduate but especially in Graduate courses.
Last but not least, i want to thank my family who support me in every way, my friends who i can count on, and especially Maria, because she gives meaning to everything that i do, and she never stops believing in me.

Abstract

In sparsely monitored basins, accurate mapping of the spatial variability of groundwater level requires the interpolation of scattered data. The methodology that is presented here is Ordinary Kriging, a method that is called the best exact interpolator, because in the absence of a nugget term, Kriging is an exact interpolator at the measurement points. In addition, Kriging allows the estimation of interpolation uncertainties which is also presented. Then, this work tackles the problem of deficient sampling of an area, due to budget constraints. To that end, the Adaptive Genetic Algorithm is being introduced, that is an Evolutionary Algorithm used for minimizing errors, and is coupled with the geo-statistical methodology to optimize the monitoring network. To do that, three different errors are defined and optimized for a constant number of measurement removals (called herein scenarios). The errors that are presented, are based either on the difference of the initial mapping with each of the reduced networks that the genetic algorithm will evaluate and evolve (RMSD, RMSE), or based on the Akaike criterion, which finds the best set of data that minimizes the error of the variogram. The described method is applied successfully to two test cases, in Mires and in Drama basin. In the first case, the initial dataset is consisted of 70 boreholes, and the method concluded that in some cases even 40 measurements could be neglected and still have an accurate mapping of the underground water level, but the safer choice would be to stop at 30 removals, because in that case, the uncertainty is much lower. Lastly, in Drama, there were 250 measurements, and the interest was to investigate the robustness of the kriging based optimization tool, and its applicability to different test cases. There, because of the bigger dataset, the RMSD was outperformed by the RMSE which only evaluates on the missing wells, instead of make the predictions in the entire grid. So a 150 removal or even 200 in some cases, where the RMSE error was more practical and Akaike was focusing more on the variogram fit. RMSD was in almost every instance slightly more accurate than RMSE except the last case when surprisingly RMSE outperformed RMSD error. So the conclusion that this work has reach is that using a genetic algorithm, and defining properly the fitness function and the succesive errors leads to a significant reduction in data measurements needed for an accurate kriging mapping. The scenario number of removals are proposed here for the two test cases, but in the end, it is a management decision of how high the uncertainty growth is allowed , or the degree of similarity of the reduced network mapping with the original dataset mapping

1 Introduction

1.1 Motivation-Physical problem and test areas description

Groundwater level monitoring networks provide essential information for water resources management, especially in areas with significant groundwater exploitation for agricultural and domestic use. Given the high maintenance costs of these networks, development of tools, which can be used by regulators for efficient network design is essential. In this work, a monitoring network optimization tool is presented. The purpose of the optimization tool is to determine which wells could be excluded from the monitoring network because they add little or no beneficial information to groundwater level mapping of the area.

The network optimization tool couples geo-statistical modeling based on the Spartan family variogram [47-51] , with a genetic algorithm method [15,19,40,41] and is applied in the first test case to Mires, located in Messara Valley in Crete Greece, an area of high socio-economic and agricultural interest, which suffers from groundwater over exploitation leading to a dramatic decrease of groundwater levels. Crete has a dry sub-humid climate and marginal groundwater resources, which are extensively used for agricultural activities and human consumption. The Messara Valley is located in the south of the Heraklion prefecture; it covers an area of 398 km^2 and is the largest and most productive valley of the island. Over exploitation during the past 30 years has led to a dramatic decrease in groundwater level, exceeding 35 m. Possible future climatic changes in the Mediterranean region, population increase, and extensive agricultural activity generate concern over the sustainability of water resources in the area and the risk of desertification. The accurate estimation of the spatial variability of the aquifer's groundwater level with the least measurements is important for integrated groundwater resources management plans. This study focuses mainly on Mires basin of the Messara Valley for two reasons. The first is the availability of hydro-geological data and the second that the basin consists entirely of alluvial sediments. Mires basin is a down-faulted trough with an area of 50.3 km^2 , roughly 14 km long and, on average, 3 km wide (Fig. 1). The trough is filled with Quaternary alluvial sediments, which form an inter-bedded sequence of gravels, gravely sands, sands, silts, silty sands and clays [14]. The data used in this study consist of 70 hydraulic head measurements, which represent averages for the wet hydrological period October–April of the hydrological year 2002–2003. The data have been provided by the Administration of Land Reclamation of the Prefecture of Crete. The measurements are unevenly distributed and mostly concentrated along the Geropotamos, a temporary river that crosses the basin (Fig. 1). The range of hydraulic heads varies from an extremely low value of 9.4 meters above sea level (m.a.s.l.) to 62 m a.s.l. for the wet period of the year. Figure 1 shows the topographic map with the locations of groundwater head measurement in Mires basin along with the corresponding surface elevation and the temporary river path as presented in [47].

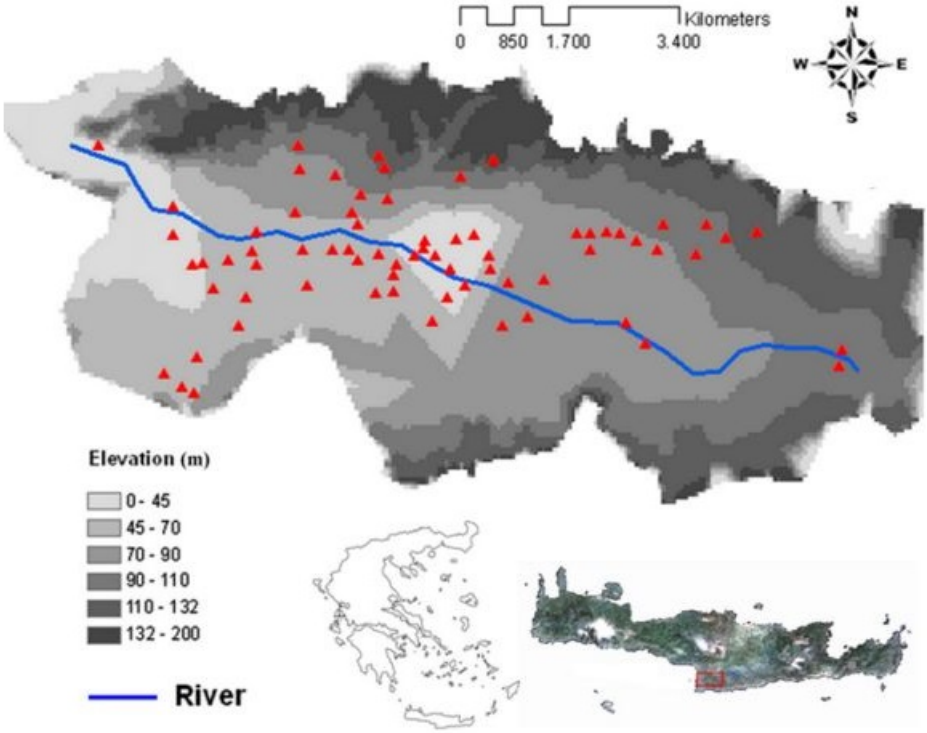


Figure 1: Main Test Case:Mires basin in Crete. Red triangles are underground measurement sites (wells)

A second test case that is considered later on this work, [44] and it is located at an alluvial aquifer at the Prefecture of Drama (Greece). The aquifer covers an area of 210 km^2 and the available data consist of 250 hydraulic head measurements that are unevenly distributed over this area. The minimum underground water level is 0.9 m.a.s.l. and the maximum is 22, so there is a range of 21.1 m, a mean value of 5.2108 m and a variance equal to 14.1420. In contrast with the Mires case where the minimum is 9.4 m, the maximum is 62 m so the range is 52.6 m, a lot bigger than Drama. As it is shown at the results, the range of the measurements has a great impact on how many wells can be excluded from the network. Lastly, in Mires the mean value of the measurements is 30.0546 and the variance equals to 153.8476. Later on this work, on the Box Cox chapter, the kurtosis and the skewness is given for the data, to check the need of the normalization transformation. The purpose of the second test case is to validate the results and assure that our robust method can be applied successfully in many test cases.

1.2 Previous work

Many works can be found in literature, that investigate the robustness of the kriging based optimization tool, and its applicability to different cases. A lot of work has been made in kriging variations for better results, for example in [47-51] there is an application of groundwater kriging in Mires basin, that compares Ordinary Kriging, [12,18,28], Universal Kriging [1,2,29,36,39], co-Kriging [18,25], Residual Kriging and Kriging with External Drift [9,16,17,52].

In [48] stochastic and deterministic methods are compared, with application in Mires. The examples that are given on deterministic methods are, inverse distance weight [5,17,34,37,43] and minimum curvature in comparison with stochastic methods, i.e. ordinary kriging (OK) [32,35,45,55], universal kriging (UK) [1,2,29,36,39] and kriging with Delaunay triangulation (DK) [20].

In [22,23,24] for the first time, Spartan family variogram method (SP) is being implemented on groundwater mapping estimation.

In [50] normalization techniques like Transgaussian Kriging, Gaussian Anamorphosis, and Box Cox [4] (Proposing a modified BC for negative data) are proposed and compared for transforming hydrological data into approximately Gaussian distributions [5,21,27,38,46]

In [15] it is presented a novel application of genetic algorithms in water level measurements in the area of Eastern Snake River Plain, developing a Kriging Based Genetic Algorithm. Although it is not with adaptive step, like our innovative work, but it is implemented for one error that is a linear weighted combination of 4 errors, including the Kriging variance, a deviation over time and a mean measurement error in addition to RMSE that is proposed in this work. The weights are chosen empirically there, so it may or may not be assumed that the optimization is truly multiobjective. In contrast with this work, where there has been presented 3 separate errors (RMSE,RMSD,AKAIKE) and explain their physical meaning and why they are to be optimized. Lastly in [15] is define the RMSD as a measure to validate their results, in contrast with this work, that RMSD is used as an error to be optimized. That is because it would have been unrealistic to optimize there in respect to RMSD because of larger datasets as in Drama's case where it will be shown that computational time increases exponentially, and is an unviable option. But on the other hand, RMSE error is more flexible and can be successfully applied in both Drama and Mires without significant loss in accuracy of the proposed reduced mapping, in comparison with the initial-data mapping.

1.3 Brief Outline

In this section, the importance of a realistic mapping is considered , and why one may expect a reduced network due to budget constraints, and then, a description of the test cases was given, and the explanation of why it is important to always have an indication of the level of the underground water there. Then, a summary of the previous work that this research was based on is given, and at the beginning of Chapter 2, a historic review of the methodology is presented. A more elaborate analysis of the two basic tools that were coupled (Geo-statistical formulation using an Ordinary Kriging with Delaunay Triangulation, and the Adaptive Integer Genetic Algorithm) is given in the main Chapters 2-3. In the fourth chapter there is a presentation on how the two algorithms were developed, and how they were modified with an innovative manner. Finally, in the last chapters, the results of our test cases in Mires and Drama, for the two different variogram methods,and for the three different errors (that are proposed at the innovation paragraph) are presented, and followed by conclusions and some future work proposals that can be made on the subject.

2 Geostatistical tools

2.1 Kriging Background-Historic review

Kriging is a geostatistical interpolation method which is known as the optimal or best linear unbiased prediction (BLUP). The French mathematician George Matheron (1963) named this method kriging, after the South African mining engineer D. G. Krige (1951), as it is still known in spatial statistics today. There, kriging served to improve the precision of predicting the concentration of gold in ore bodies [3,18]. His idea was to estimate better the gold ore grades in mining blocks, by considering the ore grades in other blocks which are close by. At the same time, Matheron had the same concern to provide the best possible estimates of mineral grades from autocorrelated sample data. He derived solutions to the problem of estimation from the fundamental theory of random processes, which in the context he called the theory of regionalized variables. From mining, geostatistics has spread into several fields of application, first into petroleum engineering, and then into subjects as diverse as hydrogeology, meteorology, soil science, agriculture, fisheries, pollution, and environmental protection. There have been numerous developments in technique, but Matheron's

thesis remains the theoretical basis of most present-day practice. [3]

In practice, in most cases the mean μ and the covariance function C(h) of the underlying random function Z(x) are unknown. Thus, unfortunately, simple kriging prediction is not applicable, since it requires information about μ and C(h) as seen in [8]. Furthermore, the most significant difference between simple and ordinary kriging, is that in the latter, the knowledge of the mean and the covariance is not assumed. For this reason, ordinary kriging represents the most common kriging method in practice and its aim is to predict the value of the random variable Z(x) at an unsampled point x_0 of a geographical region as well, as seen in [53].

Since the methodology was established, there has been an excessive use in groundwater level, with the first occurance being in [12], and other examples seen in [13-20].

2.2 Ordinary Kriging

The basic geo-statistical tool is an Ordinary Kriging method (OK) with Delaunay Triangulation [20,32,35,45,47,48,49,50,51,55] which, unlike previous investigations, uses the recently-established Spartan variogram [22,23,24,51] for groundwater level mapping. More elaborately, OK is the representation of the level of every point s_0 in our grid, by a weighted sum of our data.

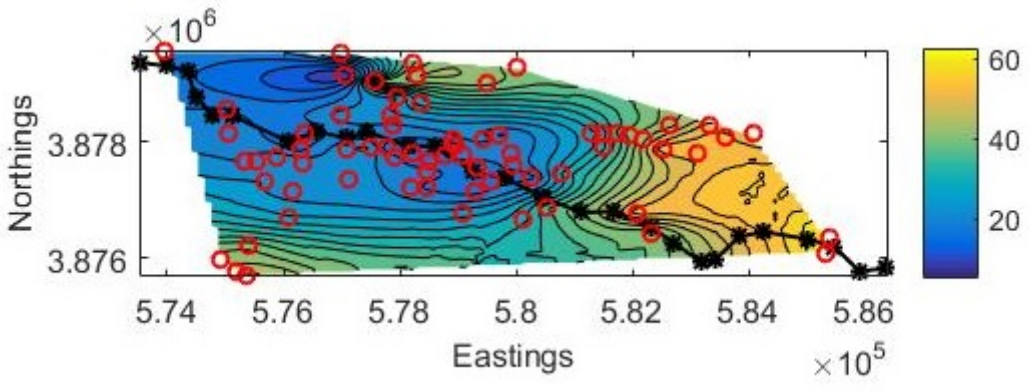


Figure 2: Mires basin OK groundwater level mapping (with Spartan variogram)

That is

$$\hat{z}(s_0) = \sum_{\{i:s_i \in \mathsf{S}_0\}} \lambda_i z(s_i) \tag{1}$$

Where S_0 is the set of sampling points in the search neighborhood of s_0.Later on this work, the question on how to select neighborhoods will arise. In our basic case study, initial data set was 70 exact coordinate locations of wells in Mires Basin, and the corresponding groundwater levels measured. The basic assumption of OK method is that z(s) is a random function with a constant but unknown mean (E[z(s)]=m). The ordinary kriging procedure is complete upon finding the corresponding weights. In order to achieve this, the interpolator is forced to

- Be Unbiased
- Minimize the kriging variance (to be defined)
- Fit the chosen theoretical variogram to Matheron Method-of-moments experimental variogram

The first bullet point translates into

$$E[\hat{z}(s_0) - z(s_0)] = 0 \tag{2}$$

which, using equation 1 gives

$$E[\sum_{\{i:s_i\in S_0\}} \lambda_i z(s_i) - z(s_0)] = 0 \Rightarrow \tag{3}$$

$$\sum_{\{i:s_i\in S_0\}} \lambda_i E[z(s_i)] - E[z(s_0)] = 0 \Rightarrow \tag{4}$$

$$(\sum_{\{i:s_i\in S_0\}} \lambda_i - 1)(m) = 0 \Rightarrow \tag{5}$$

$$\sum_{\{i:s_i\in S_0\}} \lambda_i = 1 \tag{6}$$

For the second bullet point, the errors or residuals are defined as $e(s_i) := (z(s_i) - m)$ where m is the mean value of the water level elevations. The next step is to minimize the following

$$\sigma_E^2(s_0) = E[(\hat{z}(s_0) - z(s_0))^2] = \tag{7}$$

$$E[(\sum_{\{i:s_i\in S_0\}} \lambda_i z(s_i) - z(s_0) + (m - m))^2] =$$

$$E[(\sum_{\{i:s_i\in S_0\}} \lambda_i (z(s_i) - m) - (z(s_0) - m)^2] =$$

$$E[(\sum_{\{i:s_i\in S_0\}} \lambda_i e(s_i))^2] - E[\sum_{\{i:s_i\in S_0\}} 2\lambda_i e(s_i) e(s_0)] + E[e(s_0)^2]$$

Then, using the definition of Covariance of two random variables, i.e.

$$C(X,Y) = E[XY] - E[X]E[Y] = E[(X - E[X])(Y - E[Y])] \tag{8}$$

and utilizing this into 7 it follows

$$\sigma_E^2(s_0) = \sum_{\{i:s_i\in S_0\}} \sum_{\{j:s_j\in S_0\}} \lambda_i \lambda_j C(e(s_i), e(s_j)) - 2 \sum_{\{i:s_i\in S_0\}} \lambda_i C(e(s_i), e(s_0)) + C(e(s_0), e(s_0)) \tag{9}$$

Because the second terms of the Covariance, that is $E[e(s_i)] * E[e(s_0)] = E[e(s_i)]E[e(s_j)] = E[e(s_0)]E[e(s_0)] = m^2$ cancel out because the second term has a (-2) coefficient. This is the crucial use of the basic assumption of Ordinary Kriging, that each point has an unknown but constant mean. Furthermore, 9 is needed to be minimized with respect to λ_i with the extra condition that 6 holds.In order to achieve that, one should differentiate 9 having (n+1) equations 6 with n variables. To avoid this problem, the Lagrange multiplier is inserted as the last variable in the system so that it will ensure the unbiased condition in the following way.

$$\sigma_E^2(s_0) = \sum_{\{i:s_i\in S_0\}} \sum_{\{j:s_j\in S_0\}} \lambda_i \lambda_j C(e(s_i), e(s_j)) - 2 \sum_{\{i:s_i\in S_0\}} \lambda_i C(e(s_i), e(s_0))$$

$$+ C(e(s_0), e(s_0)) + 2\mu(\sum_{\{i:s_i\in S_0\}} \lambda_i - 1) \tag{10}$$

An observation is that upon differentiating 10 in respect of μ, the unbiased condition is recovered, and upon differentiating in respect of λ_1 for example, and equal to zero to minimize, the result is

$$\sum_{\{i:s_i\in S_0\}} \lambda_i C(e(s_1), e(s_i)) + \mu = C(e(s_1), e(s_0)) \tag{11}$$

and so for every j the system of equations that one arrives is

$$\sum_{\{i:s_i\in S_0\}} \lambda_i C(e(s_j), e(s_i)) + \mu = C(e(s_j), e(s_0)) \tag{12}$$

This system in matrix notation is rewritten as

$$C * L = C_0$$

where $C = \begin{bmatrix} C(e(s_1), e(s_1)) & \dots & C(e(s_1), e(s_n)) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ C(e(s_n), e(s_1)) & \dots & C(e(s_n), e(s_n)) & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix}, L = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \mu \end{bmatrix},$

$C_0 = \begin{bmatrix} C(e(s_1), e(s_0)) \\ \vdots \\ C(e(s_n), e(s_0)) \\ 1 \end{bmatrix}$

Lemma 1 The covariance matrix C $\in \mathbb{R}^(n * n)$ of $X = (X_1, ..., X_n)$ is positive semidefinite, i.e. $\forall\ v = (v_1, ...v_n) \in \mathbb{R}^n$:

$$v^T C v = \sum_{i=1}^{n}\sum_{j=1}^{n} v_i C_{i,j} v_j \geq 0 \tag{13}$$

Proof of Lemma 1 Let a random variable $Z := \sum_{i=1}^{n} v_i X_i$ Then,

$$v^T C v = \sum_{i=1}^{n}\sum_{j=1}^{n} v_i C_{i,j} v_j = \sum_{i=1}^{n}\sum_{j=1}^{n} v_i v_j Cov(X_i, X_j) \tag{14}$$

$$= Cov(\sum_{i=1}^{n} v_i X_i, \sum_{j=1}^{n} v_j X_j) = Cov(Z, Z) = E[(Z - E[Z])^2] \geq 0$$

In this thesis it is always assumed the variance of a linear combination of random variables to be strictly positive. This assumption makes sense as in the case that the variance equals to zero, then

$$0 = Var(Z) = E[(Z - E[Z])^2]$$

and hence, because $(Z - E[Z])^2 \geq 0$, it follows that $Z = E[Z]$ almost surely for any random variable Z.

□

Using the Lemma 1, because the Covariance matrix is symmetrical, bilinear and positive semi-definite, it may be assume that it is positive definite and so the inverse matrix exists. So, the kriging weights can be found, and they are unique, and then, to find the kriging variance we multiply 12 at j step by λ_j and summing over j leads to

$$\sum_{\{i:s_i\in S_0\}} \lambda_i \sum_{\{j:s_j\in S_0\}} \lambda_j C(e(s_i), e(s_j)) = \sum_{\{i:s_i\in S_0\}} \lambda_i C(e(s_i), e(s_0)) - \sum_{\{i:s_i\in S_0\}} \lambda_i \mu \Rightarrow \tag{15}$$

$$\sum_{\{i:s_i\in S_0\}} \sum_{\{j:s_j\in S_0\}} \lambda_i \lambda_j C(e(s_i), e(s_j)) = \sum_{\{i:s_i\in S_0\}} \lambda_i C(e(s_i), e(s_0)) - \mu \tag{16}$$

Substituting this into 10 leads to

$$\sigma_E^2(s_0) = C(e(s_0), e(s_0)) + \sum_{\{i:s_i \in \mathsf{S}_0\}} \lambda_i C(e(s_i), e(s_0)) - \mu - 2 \sum_{\{i:s_i \in \mathsf{S}_0\}} \lambda_i C(e(s_i), e(s_0)) \Rightarrow \tag{17}$$

$$\sigma_E^2(s_0) = \sigma^2 - (\sum_{\{i:s_i \in \mathsf{S}_0\}} \lambda_i C(e(s_i), e(s_0)) + \mu) \tag{18}$$

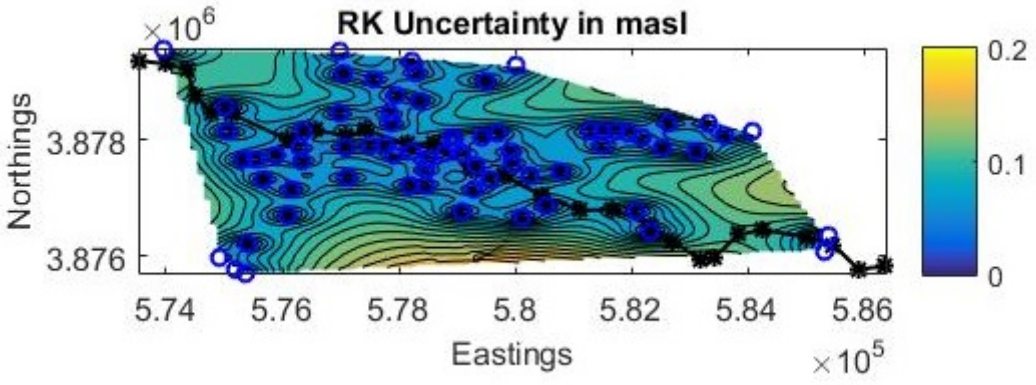


Figure 3: Kriging variance at each point of the grid with Spartan variogram in Mires

2.3 Model Variogram

The theoretical variogram is defined as half of the expectation of the squared difference of the values, so it is defined as

$$\gamma(s_i, s_j) = \frac{1}{2} E[(z(s_i) - z(s_j))^2] \tag{19}$$

$$= \frac{1}{2} E[z(s_i)^2] + \frac{1}{2} E[z(s_j)^2] - E[z(s_i) * z(s_j)] + \mu^2 - \mu^2 = E[z(s_i)^2] - \mu^2 - (E[z(s_i) * z(s_j)] - \mu^2)$$

$$\gamma(s_i, s_j) = \sigma^2 - C(e(s_i), e(s_j)) \tag{20}$$

Thus, with the help of variogram, kriging variance can be rewritten as

$$\sigma_E^2(s_0) = (\sum_{\{i:s_i \in \mathsf{S}_0\}} \lambda_i \gamma(s_i, s_0) + \mu) \tag{21}$$

According to 20, to define uniquely a theoretical variogram, it sufficed to define the covariance function between our points. So the spartan covariance function as seen in [24,48,51] is defined as

$$C_z(h) = \begin{cases} \frac{\eta_0 e^{-h\beta_2}}{2\pi\sqrt{|\eta_1^2-4|}} & for\ |\eta_1 < 2|, \sigma_z^2 = \frac{\eta_0}{2\pi\sqrt{|\eta_1^2-4|}} \\ \frac{\eta_0 e^{-h}}{8\pi} & for\ \eta_1 = 2, \sigma_z^2 = \frac{\eta_0}{8\pi} \\ \frac{\eta_0 (e^{-h\omega_1} - e^{-h\omega_2})}{4\pi(\omega_2-\omega_1)h\sqrt{|\eta_1^2-4|}} & for\ \eta_1 > 2, \sigma_z^2 = \frac{\eta_0}{4\pi\sqrt{|\eta_1^2-4|}} \end{cases} \tag{22}$$

where η_0 is the scale factor and determines in connection with η_1 the total variance of the fluctuations; η_1 is dimensionless and denotes the rigidity coefficient, $\beta_{1,2} = \frac{\sqrt{|2 \mp \eta_1|}}{2}$ $\omega_{1,2} = \sqrt{\frac{|\eta_1 \mp \Delta|}{2}}$ where $\Delta = |\eta_1^2 - 4|^{1/2}$, ξ is a characteristic length, $h = |r|/\xi$ is the normalized lag vector, where $|r|$ is the Euclidean norm and σ_z^2 is the variance.

Thanks to bibliography, there are many model variograms to fit the experimental, with the simplest being the Linear, and the two best being Spartan followed closely by the Powerlaw [22,23,24,47].

Table 1 List of theoretical semivariogram functions

Semivariogram models	
Exponential	$\gamma_z(\mathbf{r}) = \sigma_z^2 \left[1 - \exp\left(-\frac{\lvert\mathbf{r}\rvert}{\xi}\right)\right]$
Power law	$\gamma_z(\mathbf{r}) = c\,\lvert\mathbf{r}\rvert^{2H},\ 0 < H < 1$ (c is the coefficient and H the Hurst exponent)
Linear	$\gamma_z(\mathbf{r}) = c\lvert\mathbf{r}\rvert$
Matérn	$\gamma_z(\mathbf{r}) = \sigma_z^2 \left\{1 - \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{\lvert\mathbf{r}\rvert}{\xi}\right)^{\nu} \mathrm{K}_\nu\left(\frac{\lvert\mathbf{r}\rvert}{\xi}\right)\right\}$ (ν>0 is the smoothness parameter, $\Gamma(\cdot)$ is the gamma function, $\mathrm{K}_\nu(\cdot)$ is the modified Bessel function of the second kind of order ν)

Figure 4: List of theoretical semivariogram functions as seen in [47]

2.4 Algorithm of the Experimental variogram

The experimental variogram is defined utilizing the above formula of the theoretical variogram, and calculating the Expected value (E[X] of a random variable X is defined as an integral of X in respect to a probability measure, that is a bounded measure with values in [0,1], or defining a probability transition from P to P_X, which are defined from event space Ω to $\mathbb{R}$ respectively, and then we can define the expectation as $E[X] := \int_\Omega X(\omega)P(d\omega) = \int_{\mathbb{R}} xP_X(dx)$) in respect to the discrete measure with uniform probabilities $\frac{1}{N(r)}$ at the lag points (where N(r) is the number of pairs at lag r). So, the experimental variogram is defined as the average square difference of the data values between points separated by the lag vector r. So

$$\hat{\gamma}(r) = \frac{1}{2N(r)} \sum_{i=1}^{N(r)} [(z(s_i) - z(s_i + r))^2] \tag{23}$$

Here, a brief explanation of the algorithmic steps of constructing the experimental $\hat{\gamma}$ variogram is given.

- Make an array of all the possible distances, divided by the maximum distance to be normalized to [0,1]
- Select a number of lagpoints (12 was enough in our case)
- Make a n*n matrix where n is our datapoints, with all the possible distances between measurement points
- Make a n*n matrix with the corresponding differences of measurements
- Make a uniform partition of $(0, max(dist(s_i - s_j)))$ into # lagpoints (12) points
- replace each element with the mean of itself and the next one. In this way, the discretization is shifted, so that it does not include zero distances.

In this way,the data point itself is not considered in the variogram calculation. The discretization shown bellow is for example : (0.0304,0.0913,0.1522,0.2130,0.2739,0.3348,0.3957,0.4565 0.5174,0.5783,0.6391). To this shifting is owed the fact that there are 11 points calculated when # lagpoints=12 is chosen.

Due to locality effects, in this calculation it is redundant to evaluate the variogram at all distances, because a very far point is not correlated with each point of interest.

- With the previous observation, a percentage of the distances to calculate the variogram (experimental 48 %) is selected. So the new discretization is for example (0.0146,0.0438,0.0730,0.1023, 0.1315,0.1607,0.1899, 0.2191,0.2484,0.2776,0.3068)
- Find the number of points in each lag

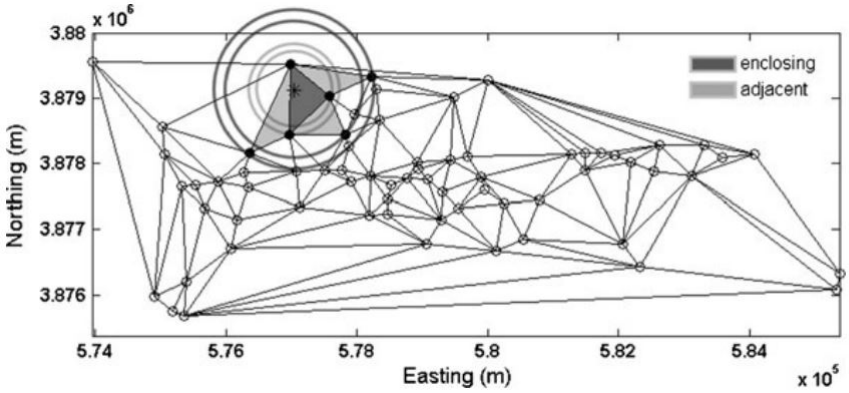


Figure 5: Example of search radius of 3 lags: (0.014,0.043) and (0.043,0.10) for one point. Here there are 3 values in the first lag and another 3 in the second. Also the relation to the Delaunay Triangulation and the first and second neighbors is shown

- Lastly, calculate the experimental variogram using formula 23

Then $\hat{\gamma}(r)$ is fitted to a model function $\gamma(r)$ with the help of the optimization Toolbox of Matlab, and function "fminsearch" which uses the Nelder-Mead Simplex Method (later on the chapter of Box Cox transform that also uses this Simplex method in its calculation, there will be more a more formal elaboration on the subject).

2.5 Analysis of the errors

In this section, the goal is to define the 3 errors that will be used to choose the best theoretical variogram method. The analysis for choosing a theoretical variogram is being done in Mires basin, because it is the main test case, and the dataset is smaller so the computations are faster. Then, the corresponding errors for the Spartan, Powerlaw and Linear, with their Variogram fitting figures are presented for the second test case in Drama, for completeness. Firstly, the Least Square Sum (LSS) is the sum of the squared difference between theoretical and experimental variogram at lagpoints, that is

$$LSS = \sum_{i=1}^{n} [\gamma(h_i) - \gamma^*(h_i)]^2 \tag{24}$$

where n is the total number of the lags, $\gamma(h_i)$ is the value of the theoretical variogram at lagpoint i , and $\gamma^*(h_i)$ is the value of the experimental variogram likewise. This is the most common error to minimize so that the theoretical variogram fits the experimental according to bibliography [10,12,16,17,47,51]

Root Mean Squared Error (RMSE) is frequently used to measure the difference between measured and predicted values of the model [47,51]. Here it is defined, after the leave-one-out process where $\hat{z}(i) :=$ "estimated values at measured locations after leaving that value out" , and we subtract it by the z(i) which is the measured value, i.e.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (\hat{z}(i) - z(i))^2}{N}} \tag{25}$$

Finally the Bias error is the same measure as the RMSE but without the 2-norm , so it can have negative values. These show that our estimates are over predicted (positive bias) or under predicted (negative bias). Unbiased estimations corresponds to zero bias.

$$BIAS = \frac{\sum_{i=1}^{N} z(i) - \hat{z}(i)}{N} \tag{26}$$

Thanks to bibliography, there are many model variograms to fit the experimental, with the simplest being the Linear, and the two best being Spartan, followed closely by the Powerlaw. [16,24,47,51]

Table 1: Comparison of variogram methods

Methods :	Spartan	Powerlaw	Linear	Exponential	Matern	Sine
LSS	$\mathbf{3.76 * 10^{-5}}$	$\mathbf{2.57 * 10^{-5}}$	$4.46 * 10^{-5}$	$6.19 * 10^{-5}$	$2.86 * 10^{-5}$	$1.26 * 10^{-4}$
time	5.0572	5.6795	**4.2155**	4.3665	13.6993	5.7924
bias	**0.2782**	**0.2915**	0.3248	0.2359	**0.2055**	0.3426
rmse	5.6450	5.6307	5.5808	5.5239	5.7711	5.5584
Akaike	**-132.4408**	**-136.6121**	-130.5492	-126.9597	-124.8402	-119.1120

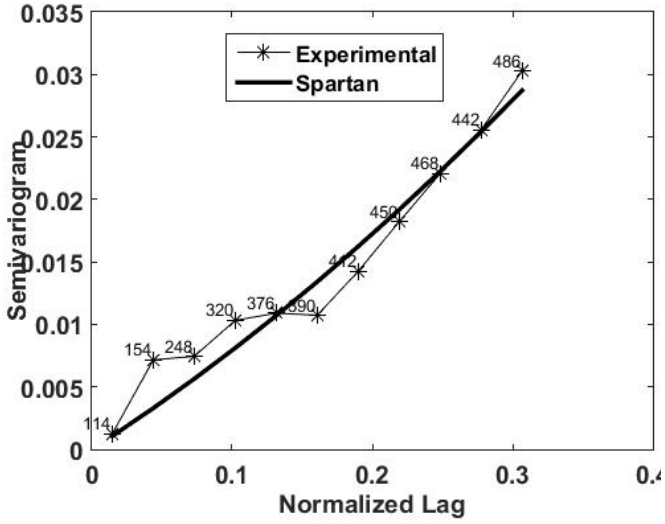
The positivity of the bias error indicates that our predictor is overestimating in every method, and of course, the lower the better applies in all 4 categories.

The criterion that is used to choose between those 6 methods was mainly because of the LSS and Akaike errors. The bias of Matern was tempting but the computational time for just one run made it non viable solution (a genetic algorithm needed approximately 10.000 evaluations in each run). Also the RMSE could be a good indicator, but it was close for all variograms as seen in the table above.

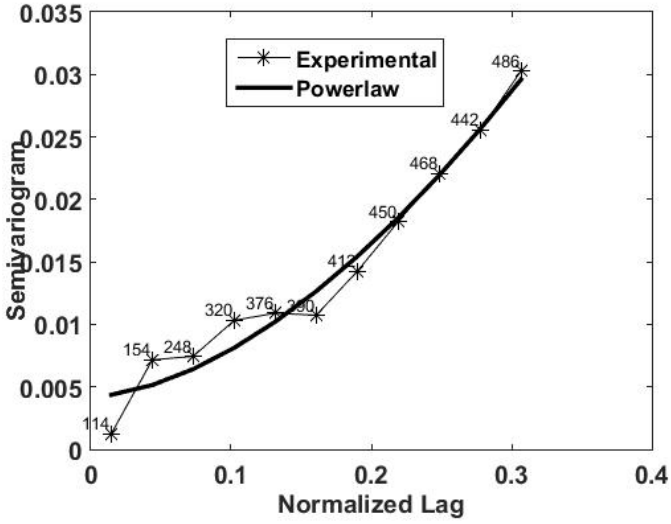
In the next table, the parameters of each variogram that needed to fit the experimental in each case can be seen for the first test case example, where η_1 is the rigidity coefficient at Spartan, H is the Hurst exponent at Powerlaw, which is a number $\in (0, 1)$, and the Matern variogram needs a ν parameter which is a smoothness parameter as seen in [87–89].

Table 2: The Optimized parameters for our Model Variograms in Mires basin

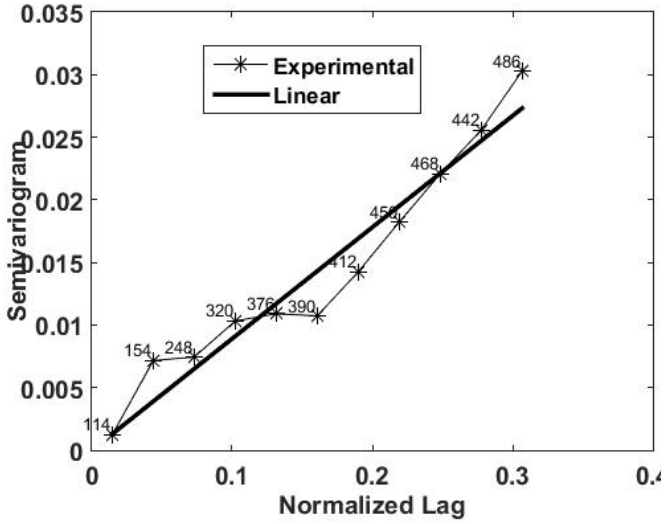
Methods:	Spartan	Power	Linear	Exponential	matern	sine
σ_z^2 or c	0.2507	0.1903	0.0890	0.0718	0.1519	0.0553
ξ	0.4905	N/A	N/A	0.6625	1.0205	0.1483
nugget	N/A	0.0042	N/A	0.0006	0.0044	N/A
Other parameters	$\eta_1 = -1.9799$	H=1.7052	N/A	N/A	smoothness $\nu = 1.5714$	N/A



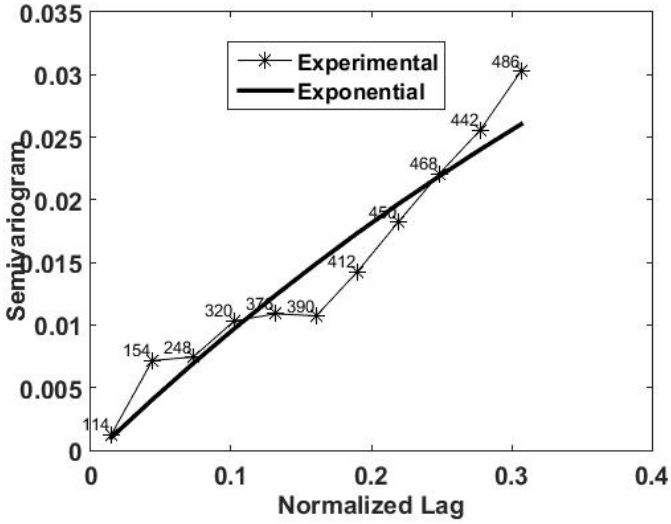
(a) spartan variogram



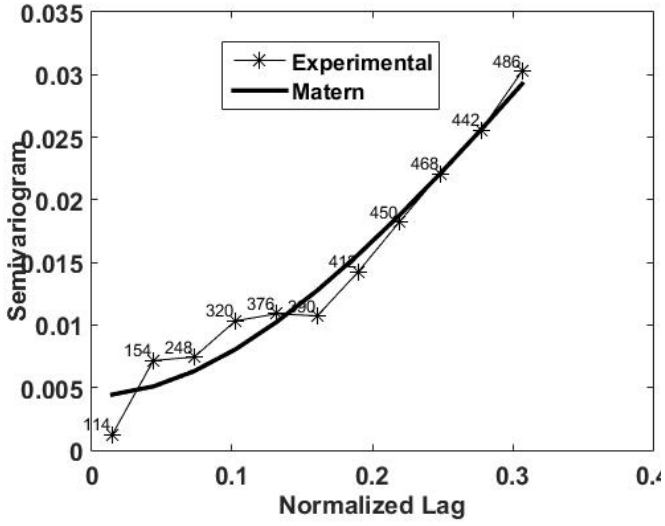
(b) powerlaw variogram



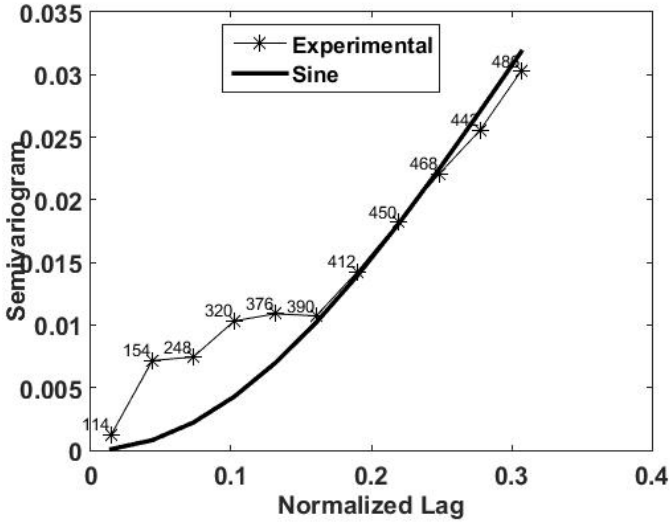
(c) linear variogram



(d) exponential variogram



(e) matern variogram



(f) sine variogram

Figure 6: Theoretical variogram methods examined in Mires basin

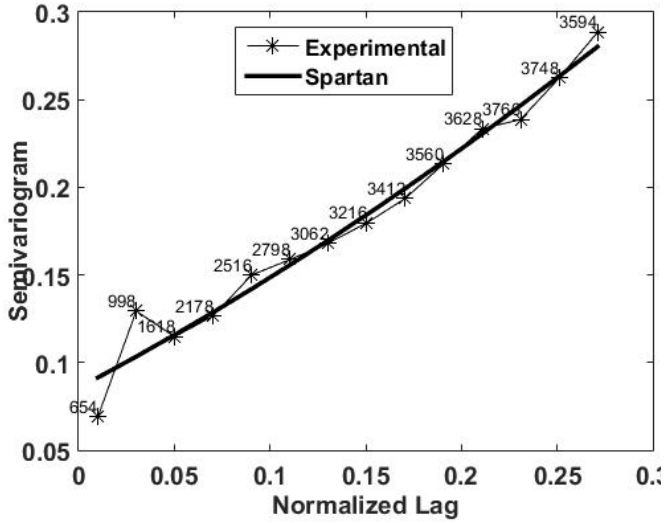
Lastly in this subsection, as it is mentioned in the prologue, the errors of Spartan, Powerlaw and Linear in Drama are presented.

Table 3: Comparison of 2 best variogram methods and Linear in Drama

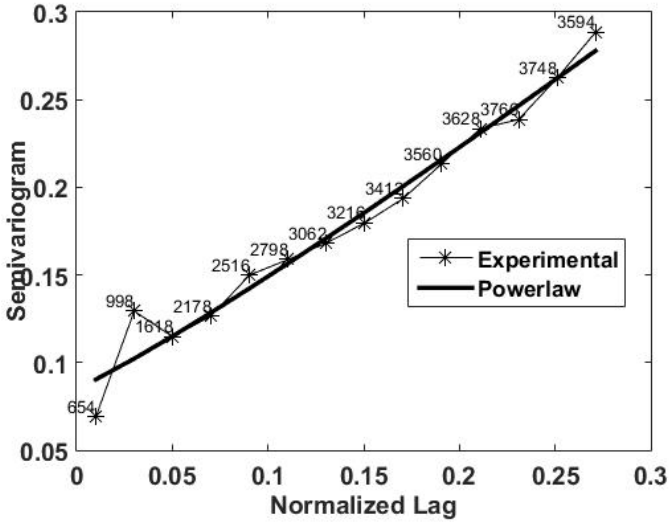
Methods :	Spartan	Powerlaw	Linear
LSS	**0.0015**	**0.0015**	0.0016
time	24.7105	40.5407	19.0111
bias	0.3474	**0.3319**	0.3531
rmse	**2.8394**	2.8856	2.8406
Akaike	-120.4037	-121.6538	-123.2640

Table 4: The Optimized parameters for our Model Variograms in Drama

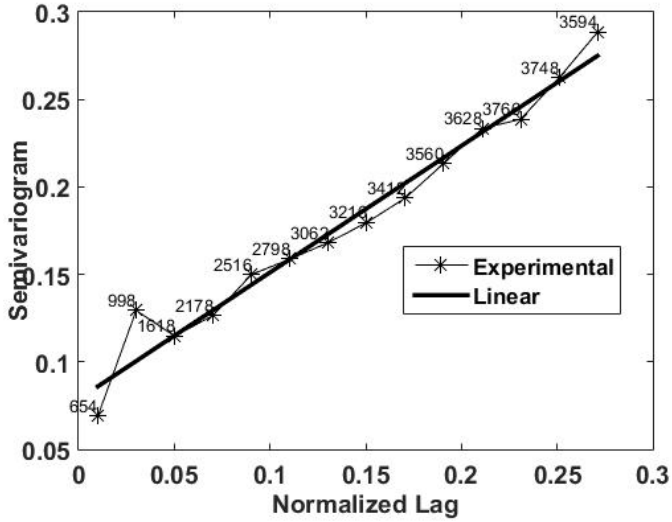
Methods:	Spartan	Power	Linear
σ_z^2 or c	1.8053	0.8175	0.7238
ξ	0.9294	N/A	N/A
nugget	0.0856	0.0854	0.0787
Other parameters	$\eta_1 = -1.8929$	H= 1.1078	N/A



(a) spartan variogram



(b) powerlaw variogram



(c) linear variogram

Figure 7: Theoretical variogram methods examined in Drama basin

As observed from the figures and the tables above, spartan and powerlaw variograms are very similar, with a slightly better rmse in spartan, and a better bias in powerlaw. Linear is more simplistic in its calculations, so it is faster, and is presented only for comparison purposes. Herein, in both areas, the results are presented with Spartan and Powerlaw variogram methods.

2.6 Delaunay Triangulation

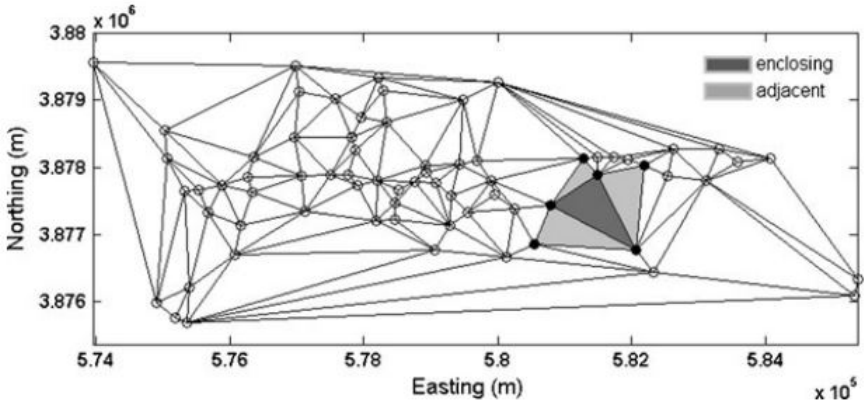


Figure 8: Delaunay Triangulation at Mires basin. First neighbours shown

Kriging weights at locations with low correlation would be very small, so, to avoid computing extra values that have little to no difference to our result, our neighbourhoods are chosen by making a Delaunay Triangulation as seen in the figure above, and include for each point in our grid the first and the second neighbouring triangles. Also a very useful application of this methodology, is that, from the full grid (rectangle 100x100 points), only the ones inside the hull that is created by the measurements have a predicted value (interpolation and not extrapolation). That is the most that is expected, due to physical problem constraints, a point outside the hull can not be predicted with precision as only one or two points will be its neighbours.

2.7 Box Cox Tranformation

The Box-Cox (BC) method (Box and Cox 1964) is widely used to transform hydrological data into approximately normal distributions as seen in [4,49]. That means that it optimizes the k exponent as it can be seen in the definition to make the kurtosis of the data near to 3 and the skewness near to 0. The transform is defined only for positive data [4] values and is defined by means of:

$$y = g_{BC}(z;k) = \begin{cases} \frac{z^k - 1}{k} & k \neq 0 \\ log(z) & k = 0 \end{cases}$$

Where skewness is as known defined to be

$$S = \frac{\sum_{i=1}^{N}(Z_i - m)^3}{N * \sigma^3} \tag{27}$$

and kurtosis

$$K = \frac{\sum_{i=1}^{N}(Z_i - m)^4}{N * \sigma^4} \tag{28}$$

Where m is the mean value of the observations Zi and σ is the standard deviation. The skewness for a normal distribution is zero, and any symmetric data should have a skewness near zero. Negative values for the skewness indicate data that are skewed left and positive values for the skewness

indicate data that are skewed right. Accordingly, the kurtosis for the normal distribution is 3 and if kurtosis is more than 3 indicates a "heavy-tailed" distribution and if it is lower than 3 it indicates a "light tailed" distribution.

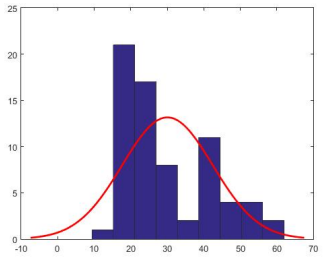
Given the vector of data observations $z = (z_1, ..., z_N)$,the procedure of the transform is to find the optimal value of the power exponent k that leads to the best agreement of $y = (g_k(z_1), ..., g_k(z_N))$ with the normal distribution.

Having found the correct exponent $k \neq 0$ to transform the data, after the OK procedure is over, the correct scale of the levels can be recovered by solving the definition of the transform for z and apply it to the predictions, that is

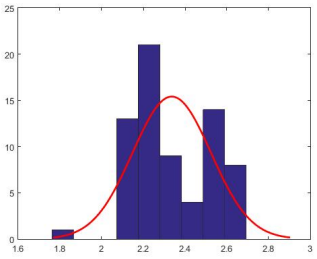
$$z = (k * y + 1)^{1/k} \tag{29}$$

where y is now the predictions under the transformation, while z is the predictions on the initial scaling.

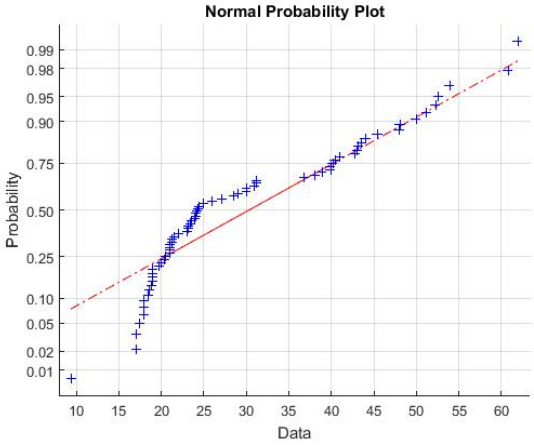
In Mires basin, initially the data had
K=2.5848
S=0.8140
and after the transform, an exponent is found to k=-0.2239,
and it transforms the data and gives a new K=2.7037 and S=0.0120, obviously closer to 3 and 0 respectively.



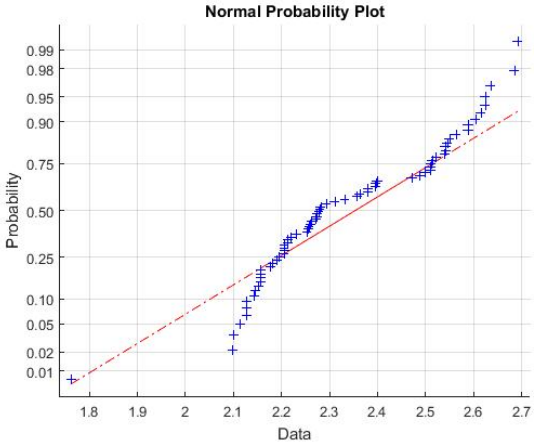
(a) Histogram fit before BC



(b) Histogram fit after BC

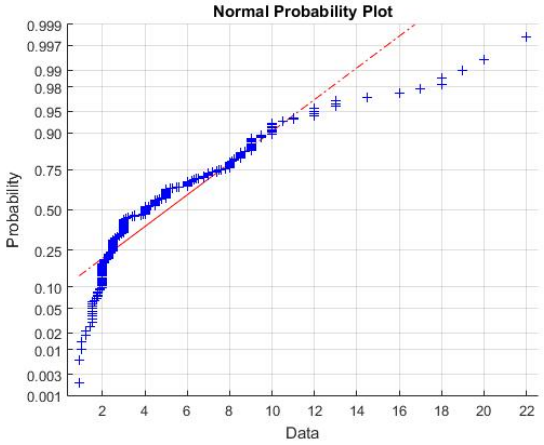


(a) normal distribution plot before BC

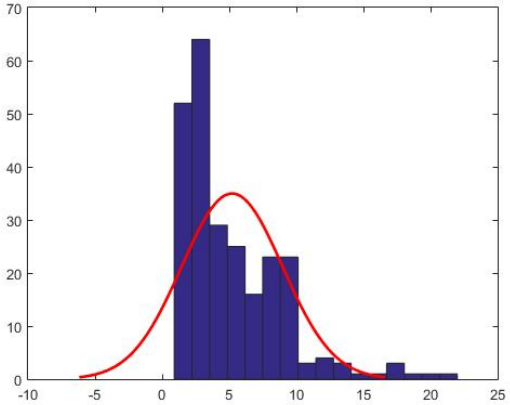


(b) normal distribution plot after BC

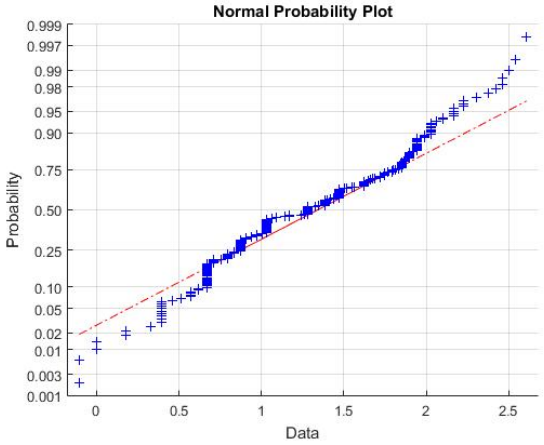
The same procedure is followed also in Drama, where the transformation is more crucial because the initial data had kurtosis
K= 6.0452
and skewness
S= 1.5839
and with the help of the Box cox transform, the resulting values are
K=2.2230
and S= 0.0201
which are much closer to 3 and 0 respectively. Below, the normal distribution plot and the histogram fitting to the normal distribution are shown for the Drama test case before and after the Box Cox Transform (BC).



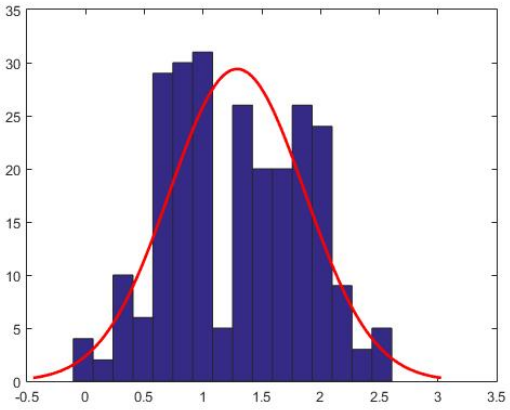
(a) normal distribution plot before BC in Drama



(b) Histogram fit plot before BC in Drama

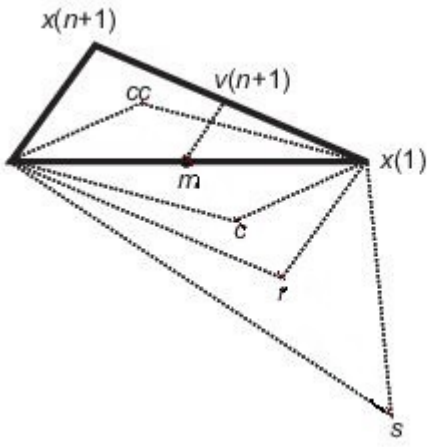


(a) normal distribution plot after BC in Drama



(b) Histogram fit plot after BC in Drama

This operation is determined with the function boxcox which is a built in function of Matlab and it uses "fminsearch" minimization which can be used for unconstrained multivariate problems, and is a derivative-free minimization technique. More elaborately, "fminsearch" uses the Nelder-Mead simplex algorithm.This is a direct search method that does not use numerical or analytic gradients. This algorithm uses a simplex of n + 1 points for n-dimensional vectors x. It computes the values of the function at these vertices and then, the algorithm modifies the simplex repeatedly with 3 basic procedures on the worst vertex, comparing with the worst value of the simplex. If it is better it creates a new simplex an so on. The above procedure is presented in the simple case of a two dimensional simplex (triangle) in the following figure.



Nielder-Mead simplex Algorithm

1)calculates biggest value at x(n+1) and lowest at x(1)

2)reflects the point at direction of m to r and calculates f(r)

3)If f(r)<f(x(1)) expand it to s

4)If f(r)>f(x(n)) do a contraction to c or cc acordingly

5) keep x(1) and shrink all the other sides to shrink the simplex (x(n))'=x(1)+(x(n)-x(1))/2

Figure 13: Simplex algorithm formulation

3 Genetic Algorithms

As seen in the textbooks [40,54] the genetic algorithm method, that belongs to the family of evolutionary algorithms, is a near optimal technique, meaning that, given time, the algorithm will find a near minimal value of a fitness function. This is important in cases in which we do not need the accuracy of "the best scenario", in highly non linear functions, and non differentiable, that other minimization techniques can not be applied. The obvious drawback is that one can never be sure that the optimum value has been reached, and if not a suitable termination criterion has been set, then the algorithm can even return a local minimum which is misleading. As quiting from the textbooks there are 5 basic processes that a genetic algorithm goes through, that is

- Creation
- Evaluation
- Elitism
- Crossover
- Mutation

3.1 Creation

In the first process, after a population size is selected, a random population of that size is being created. There are some options, that can contain the initial population in some space of our choosing, that may be the set on which the fitness function can be defined.

3.2 Evaluation

In the second process,each and every member of our population is evaluated, using the fitness function. Then, the population is rearranged based on each fitness value from the better(lowest) to the worst(higher). This initial population is called parents.The next population that will be created by the parents will be called children or kids, and so a generation passes. Then the children are named parents and they create kids of their own and so on.

3.3 Elitism

There is a small number of parents in the amount of

$$\max(1, (0.05 * PopulationSize))$$

at the top of the population that is named the Elitistic kid or kids , and it will stay unchanged. Elitism protect the best candidate(s) of each population, to stay in the next population.

3.4 Crossover

Crossover is the process where our population is mixed to produce better fitness, so usually a big fraction of our population is used for crossover.The ways to make a crossover between parents are plenty, to name a few, One point crossover, Two point crossover, Uniform crossover etc. To understand these better in our case, 1-point crossover is seen in the following example, where one genome (here the first) is being chosen randomly and it is traded from one parent to another.

$$\begin{array}{rcccc} parent_1: & 1 & 3 & 5 & 7 \\ parent_2: & 0 & 2 & 4 & 6 \\ child_1(1point) & 0 & 3 & 5 & 7 \\ child_2(1point) & 1 & 2 & 4 & 6 \end{array} \tag{30}$$

The more interesting type of crossover in this context is the Uniform type of crossovers where there is a crossover rate (0.5 at uniform) at which each genome is traded. An example that applies to our case is being presented shortly, where if a 5-removals-scenario is assumed, and the initial data-wells are numbered from 1 to 70, then with .5 crossover rate the result is:

$$\begin{array}{rcccc} parent_1: & 10 & 35 & 54 & 69 \\ parent_2: & 11 & 52 & 41 & 6 \\ child_1(uniform) & 10 & 35 & 41 & 6 \\ child_2(uniform) & 11 & 52 & 54 & 69 \end{array} \tag{31}$$

3.5 Mutation

Similarly to crossover, mutation can be implemented with several methods, like one-point, uniform, gaussian etc. The example of one point and uniform mutation is given, resulting in

$$\begin{array}{rcccc} parent: & 1 & 3 & 5 & 7 \\ child_1(onepoint) & 1 & 2 & 5 & 7 \end{array} \tag{32}$$

Here one genome is changed randomly from the values of the search space. Addressing the problem at hand, if a 5-removals-scenario is assumed again, then a mutation would change randomly 2 of 4 genomes at a random value in the search space.

$$\begin{array}{rcccc} parent: & 10 & 66 & 16 & 17 \\ child_1(uniform) & 10 & 66 & 5 & 7 \end{array} \tag{33}$$

Crossover-Mutation fraction A number between (0,1) is defined to be the Crossover fraction, that is the percentage of the rest of the population after elitistic kids have been removed that will undergo crossover, and the rest of it will undergo mutation. So

$$CrossoverFraction = 1 - \frac{EliteKids}{PopulationSize} - MutationFraction$$

Because the crossover is the most important way to reproduce (called **exploitation**), there is a higher fraction for it (usually 0.8) so , minus the elitistic children , there is a less fraction to mutate. Mutation is the **exploration** on the search space, meaning that it is the random part of the evolution that is much needed because of the fact that the optimal value that is sought, may not be accessible through mixing our initial population.

3.6 Termination Criteria

There are plenty termination criteria that have been proposed in the bibliography, to name a few, there is Time criterion where user has to set a time limit, there is Generations criterion, where a total generation limit is set, there is Tolerance in Best fitness value or Mean fitness value where the user sets beforehand a small tolerance number, and when the difference between mean fitness value or best fitness value is below that tolerance the algorithm stops. In this research, as mentioned before, a stall generations criterion have been selected for termination, which is one of the most reliable, in which you allow the generic algorithm to evolve to as many generations as it needs, and stop only if it has unchanged best fitness value for # Stall Generations. To decide what population size and stall generation to choose,a sensitivity analysis will be presented separately for the 3 different statistical measures that were optimized, and that will be defined in the next section.

4 Model development-innovation

4.1 Innovation to Standard Genetic Algorithms

One of the modifications to classic genetic algorithm that has been done in this work is, that through the generations, the crossover-mutation fraction does not remain constant. It is programmed to adaptively change according to the optimization needs. For that purpose the Stall parameter is introduced, that is defined to be the number of generations that has passed without a change at the best fitness value. That means that the algorithm may have found the optimum value, but most of the times means that it can not produce better children by exploitation, and so it needs to do more exploration. So, the criterion is that for every 10 stall generations passed, we add 0.10 to the mutation fraction and lower 0.10 the crossover fraction. [41] By the same criterion, the crossover and mutation rate is also changed, and this results to a change in exploration and exploitation as well. Next. because of the integer search that is being implemented (instead of usual search over the Reals) , and because of the lack of Matlab (2016a) pre-programmed functions that can adapt as described over the integers, the genetic algorithm optimization tool of Matlab was customized according to the needs, by creating the functions of Creation of initial population, Crossover and Mutation. One of the biggest drawbacks using a genetic algorithm is the time that it needs to conclude, and the fact that if one wants great degree of accuracy, a great deal of computational time needs to be sacrificed. For that manner, to make the operation faster, a vectorized form was set, meaning that in every generation, the fitness function is evaluated only one time, and gets as input the entire population as a vector. Lastly, the modification that has been made in the geo-statistical tool to optimize the speed of the genetic algorithm, was that there were made two separate programs for the spartan and power law variograms, because evaluating in the full grid, and make an if statement 100x100 times slowed the system down. Also the localization technique using the Delaunay that has been described above, also contributed to a faster procedure.

4.2 The errors for the optimization

Three methods of optimization were used, the two of them working only by making an estimate on the missing wells, so they are very fast to implement, but lacks accuracy as often happens with speed and accuracy. The first (Root Mean Squared Error-RMSE) is a wide known measure for cross validation purposes as seen in [28,48,51,53,55]. The cross validation process is, as seen in a previous chapter, a procedure where one by one, all of the data are subtracted, and an estimation is given using the rest of the data. By this inspiration, the optimized RMSE here is calculated by subtracting # scenario wells, and estimate at the missing wells using the rest of the network.

So,

$$E_{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}[\hat{z}(s_i) - z(s_i)]^2} \tag{34}$$

Where N is the number of wells that have been left out, $\hat{z}(s_i)$ is the estimated level using the rest of the (n-N) data, and $z(s_i)$ is the measured level at the corresponding measurement site that has been neglected.

The second method that will be described is called Akaike Criterion as seen in [44], and it is also calculated at the cross validation stage, so it is also very fast implemented by the genetic algorithm as it will be presented later on. Akaike error is defined as

$$E_{AIC} = nlog[\frac{\sum_{i=1}^{n}[\hat{\gamma}(h_i) - \gamma(h_i)]^2}{n}] + 2\mu \tag{35}$$

where n is the lag number, $\hat{\gamma}(h_i)$ is the value of the experimental variogram at lag i, $\gamma(h_i)$ is the value of the theoretical variogram at lag i and μ is the number of degrees of freedom that each theoretical variogram has. This criterion is mainly used to compare each method of variograms with each other (obviously the lower the better) and it also takes into account how many parameters each variogram has. Minimizing with this measure not only gives an excellent fit to our variogram (because despite the fminsearch, it also chooses the data to make the perfect fit) but also it gives an indicator of which is the best method that works better with the reduced network. Last but not least, the resemblance to the LSS presented above is notable, but in for the optimization purposes, the Akaike criterion was preferred.

The third method is inspired by [15] where it is used as a method to check their results. It gives the most accurate mapping in comparison to the original mapping, and is defined to be the Root mean squared difference between the 2 matrices (symbolized RMSD -root mean squared deviation). So

$$E_{RMSD} = \sqrt{\frac{1}{M}\sum_{i=1}^{M}[z_{init}(s_i) - \hat{z}(s_i)]^2} \tag{36}$$

Where M is the number of points inside our triangulation in our 100x100 grid, z_{init} is the predicted values at the M points using the original network and $\hat{z}$ is the predicted values at the same M points using the reduced network.

4.3 Matlab algorithm formulation

All the scenario experiments were run using an Intel i5 6600 processor at 3.3 Ghertz with 8 Gb RAM with the use of MATLAB 2016a version.

For the coupling of the two tools mentioned above (geostatistics and genetic algorithms) there were two basic programs on matlab which took as input a vectorized set of populations' length, which consist of the scenarios' length of possible combinations of wells to exclude. For example, if population size was set to 100, and a 30 scenario exclusion was assumed, in each generation the programs automated input from the genetic algorithm would be a 100 times [1x30]. In the vectors 1x30, the constraint that each number did not appear twice is forced, and each well number is in the margin of [1,70] in natural numbers. (For the first test case). So, the creation function of the standard genetic algorithm has been customized, so that it does all the mentioned above actions, creating the initial population. Next, the second customization that has been made, was creating custom mutation and crossover functions, because an integer optimization was required, so values had to be interchanged from 2 parents with a crossover-mutation rate (adaptivelly changed), and each time have a constraint so that each child is a feasible solution, meaning that it did not have a same well twice (ensure a 30-and-no-less well scenario) and each candidate have well numbers from 1 to 70. Mutation is also customized by the same manner, with random selection in mutation rate % of the parent's length. The percentage of the population to be crossovered or mutated (as mentioned crossover-mutation fraction) is also changed adaptivelly, and is programmed into the genetic algorithm function of Matlab, which was changed accordingly.

Finally, for the fitness function formulation, there were 2 functions of each variogram method (Spartan and Powerlaw). For the faster errors RMSE and Akaike, the only program that had to be included in the fitness function was the data loading, the boxcox transform and triangulation, and then the variogram estimation (where the LSS is estimated and the Akaike).Lastly, the estimation of the excluded wells using the remaining reduced network of wells is being made. For the minimization of the RMSD error, which is more time consuming, a more sophisticated algorithm had to be done, in which the whole computation of kriging in all the points in our 100x100 grid had to be evaluated, inside the triangulation domain. Also the estimation of the uncertainty that comes with the ordinary kriging computation is being made, and an observation is that, as expected, having removed more and more measurements makes the mapping more and more uncertain, but in low exclusion scenarios, providing that the area initially was oversampled, there is not significant increase in uncertainty. That is also a measure to find out in which level we can stop excluding more and more wells (In Mires it was about 40 out of 70 and in Drama was 150 or in some cases even 200 out of 250).

4.4 The necessity of a genetic algorithm tool

To excess the need of a genetic algorithm to find a local minimum, bellow are shown the 3 errors for only one removal, how they behave and how irregular the pattern is. (moErr is the RMSD and moPR is the RMSV)

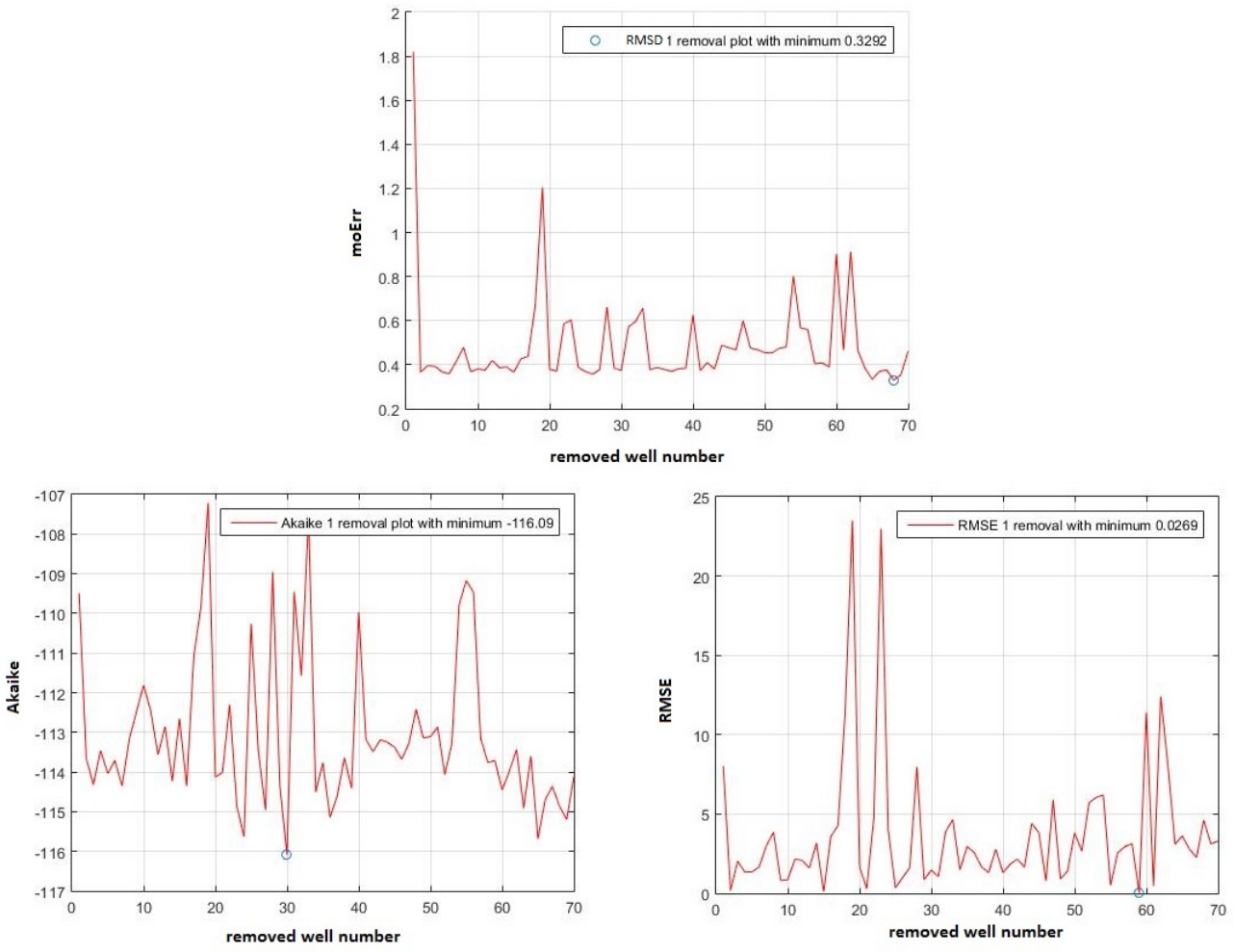


Figure 14: 3-error comparison for 1-removal scenario with their minima

Of course for the 30 removal scenario in Mires that is proposed in this study, there are

$$\binom{70}{30} = \frac{70!}{30! * 40!} \approx 5.5348 * 10^{19} \tag{37}$$

and if a hard search by the means of RMSD that needs approximately 2 seconds to run each time was being sought, we would need approximately

$$11.0696 * 10^{19} sec \approx 5930 * 10^{8} years$$

to complete.

4.5 Sensitivity analysis of adaptive G.A.

In this subsection, a sensitivity analysis is being made in Mires basin, with the standard 30 removal scenario, at RMSE error,to check the sensitivity of the adaptive change in crossover-mutation fraction and crossover and mutation ratio. In the basic G.A., a 0.85 crossover fraction is being used, standardized so the remaining 0.15 was mutation minus the elitistic children which are the standard 0.05* Population size. In the adaptive case it is proposed to start with a 0.85 crossover fraction and in every 5 Stall generation increase, the percentage is dropped by 10 percent. There is also a semi adaptive case, in which crossover fraction begins with 0.85 and it drops to 0.5 when stall is

over 10. The results are presented below at a micro genetic algorithm (meaning a little population and stall just for benchmarking purposes) with 50 population and 20 stall

Table 5: Adaptivity effect

	Generations	Fvals	RMSE
Adaptive	118	5950	8.042
Semi adaptive	103	5200	9.004
Standard	94	4750	10.27

As it can be validated from our results, even with such a small population of 50 candidates in each generation, there is a clear advantage as the crossover fraction reduces when the stall generations increases, and that is because it usually can not reach better solutions with crossover when the stall is high, and it needs more randomness that is achieved by mutating more.

5 Test case 1: Mires Basin

The basic case study was in Mires, an area of high socio-economic interest as mentioned before. As the results indicates, from the initial 70 well network, it can be easily deduced that 30 can be removed and keep the quality of the mapping, and in some cases, even 40 measurements can be neglected. Of course it can be observed, if the goal is to try to leave the mapping unchanged (RMSD, RMSE), the Kriging Variance will rise, because the uncertainty will rise. Lastly, when an optimization with respect to Akaike error is being implemented, it is observed that there is an almost perfect interpolation between the theoretic and experimental variogram and a relative smoothening in the change of values at the predicted mapping.
Before the figures,a table with the errors which are minimized is demonstrated, and the deviation from the initial mapping, when RMSD is not optimized genetically (This is useful to know if the proposed mapping is similar to the initial mapping.)

Table 6: SPARTAN OPTIMIZATION

Scenarios :	30	40	50
RMSD optimization	0.5671	0.8250	1.3542
RMSE optimization	6.8550	14.2811	17.9336
Corresponding RMSDs	0.9528	0.9621	1.1672
Akaike optimization	-158.3399	-165.3199	-192.8193
Corresponding RMSDs	2.2906	4.9930	8.6190

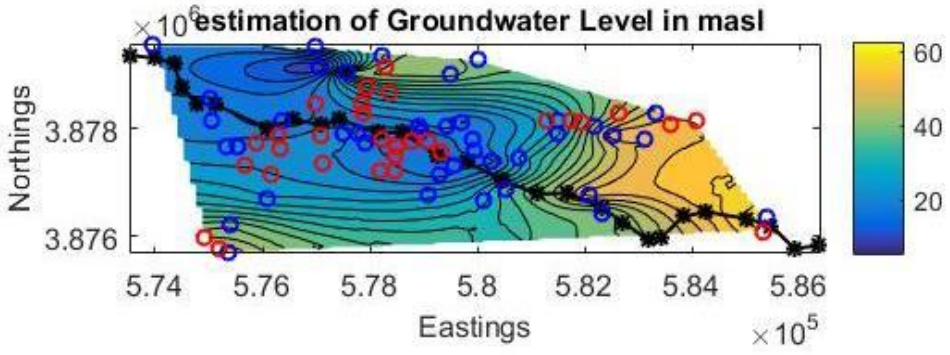
Table 7: POWERLAW OPTIMIZATION

Scenarios :	30	40	50
RMSD optimization	0.8693	1.0608	1.4854
RMSE optimization	6.6399	10.8327	18.7995
Corresponding RMSDs	0.9169	1.3339	2.2506
Akaike optimization	-163.3718	-160.4464	-186.1825
Corresponding RMSDs	5.1526	4.3471	10.4693

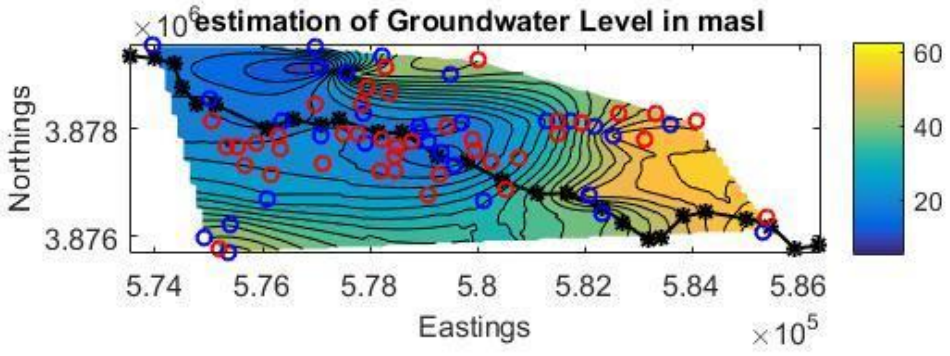
It is easily observed that Spartan is better at almost every instance, and that, as expected, as there are more subtractions from the original dataset, the RMSD,RMSE errors are keep growing and the Akaike finds new minimal, and that is due to the fact that it is dependent only on variogram fitting, and it is easier to fit a curve to less data. But as one may observe, the corresponding RMSD of Akaike is non viable, and that fact shows in the next figures where these results are presented.

5.1 Spartan Optimization

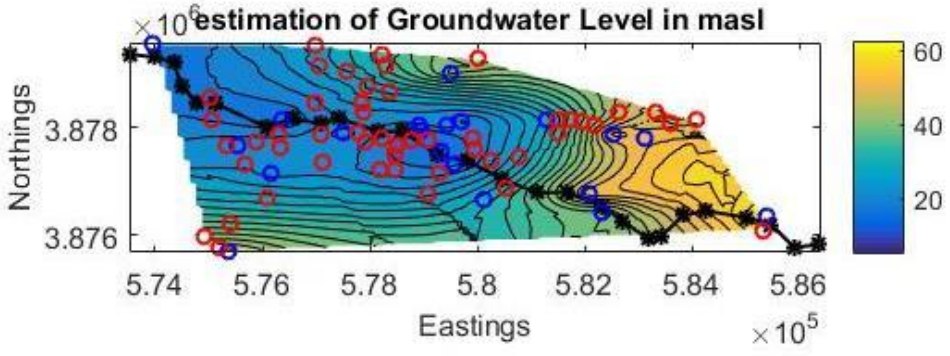
The results for the Spartan (SP) optimization in Mires are presented in this section, where a 30,40 and at some cases 50 data points from our 70-data initial mapping were able to get subtracted without significant degradation of the mapping.



(a) Spartan RMSD 30

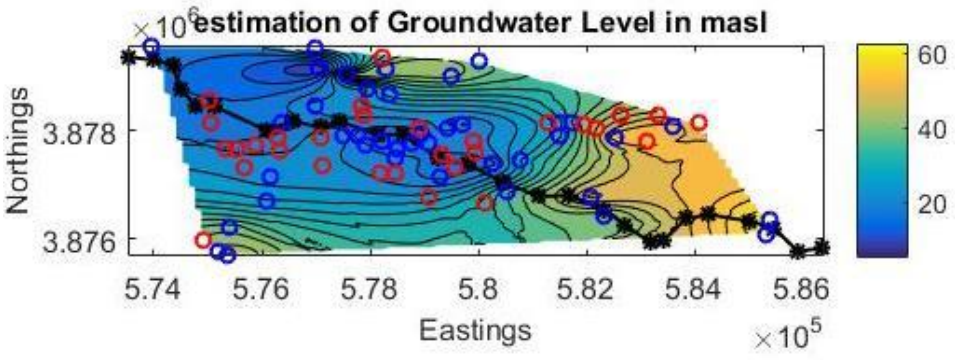


(b) spartan RMSD 40

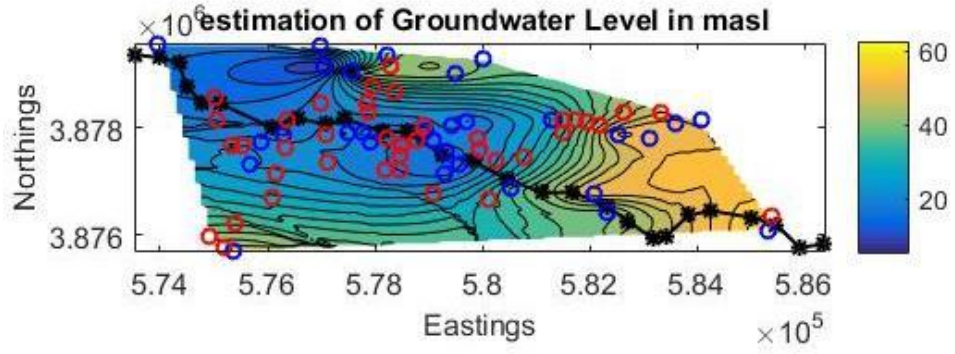


(c) spartan RMSD 50

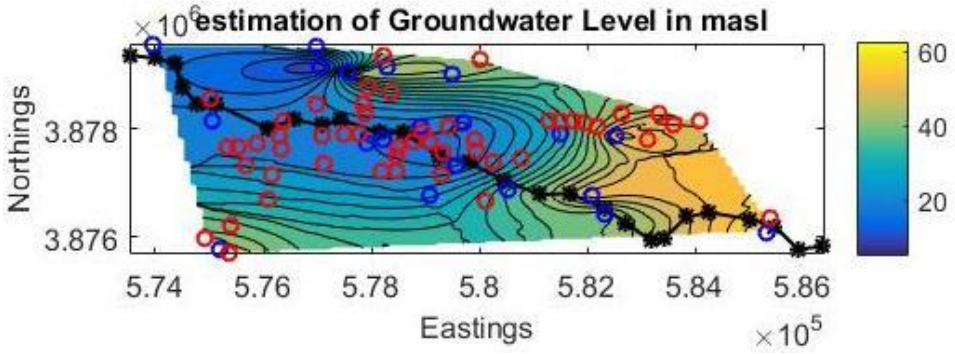
Figure 15: RMSD minimization scenarios mapping



(a) Spartan RMSE 30



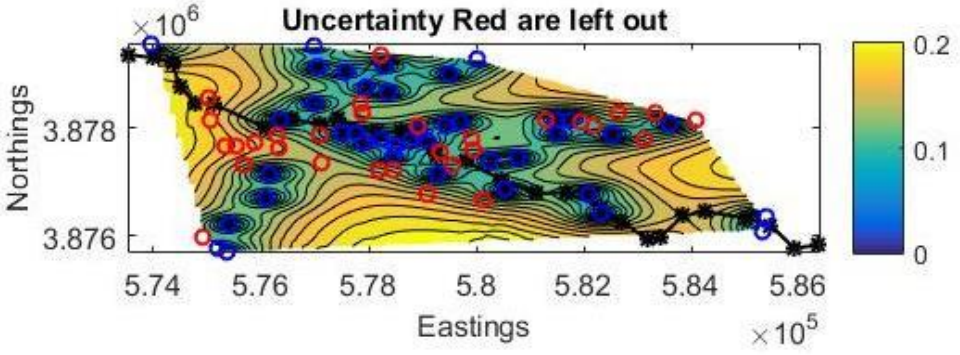
(b) spartan RMSE 40



(c) spartan RMSE 50

Figure 16: RMSE minimization scenarios mapping

At both RMSD and RMSE errors minimization there is a close resemblance with figure 2 that is the initial mapping in Mires with SP variogram. This was expected since that is what these errors represent (difference from the original mapping). Surprise is the fact that in some cases, RMSE (which is faster) is more accurate than RMSD (The fact that RMSE is faster, allowed us to tune up the Population size and stall limit critirion, so that it had time to explore and exploit more.



(a) spartan RMSE uncertainty 30

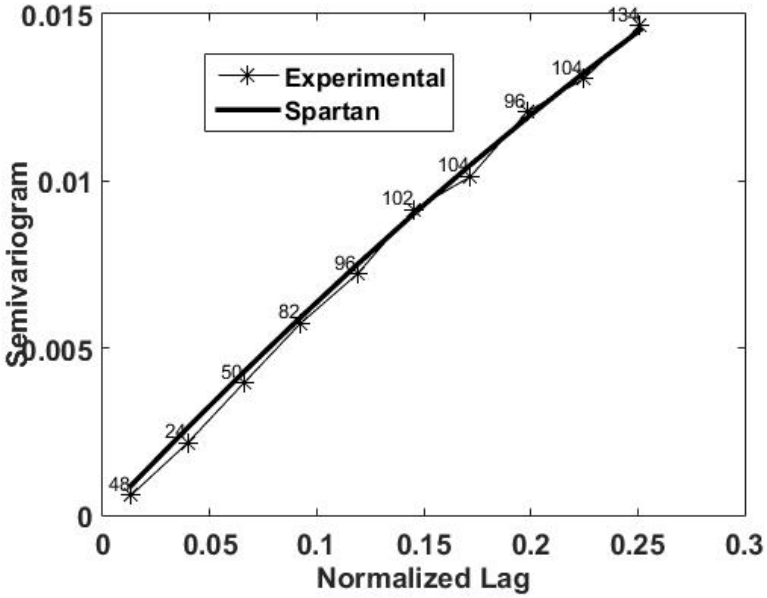
(b) spartan RMSE uncertainty 40

(c) spartan RMSE uncertainty 50

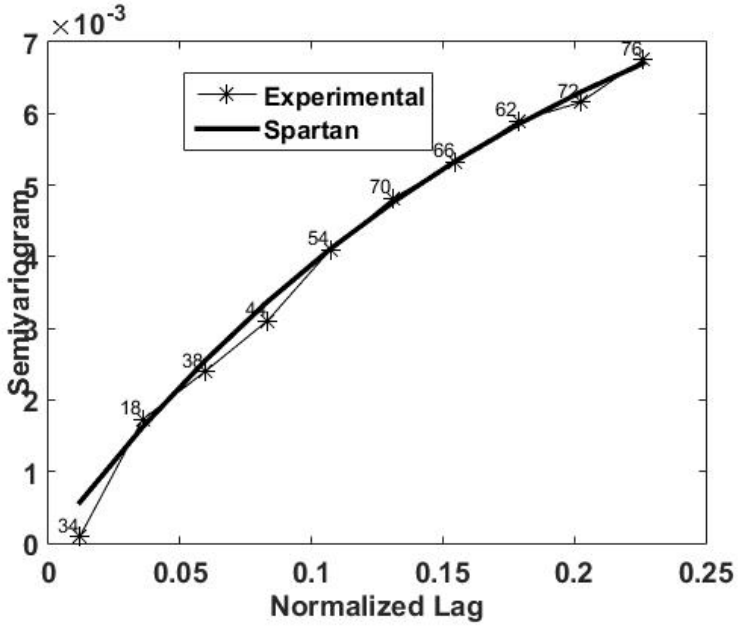
Figure 17: RMSE minimization corresponding uncertainties

In the above three figures it is shown how the uncertainty (kriging variance) degraded over the entire grid, and that the more wells are excluded from the network, the more the uncertainty grows.

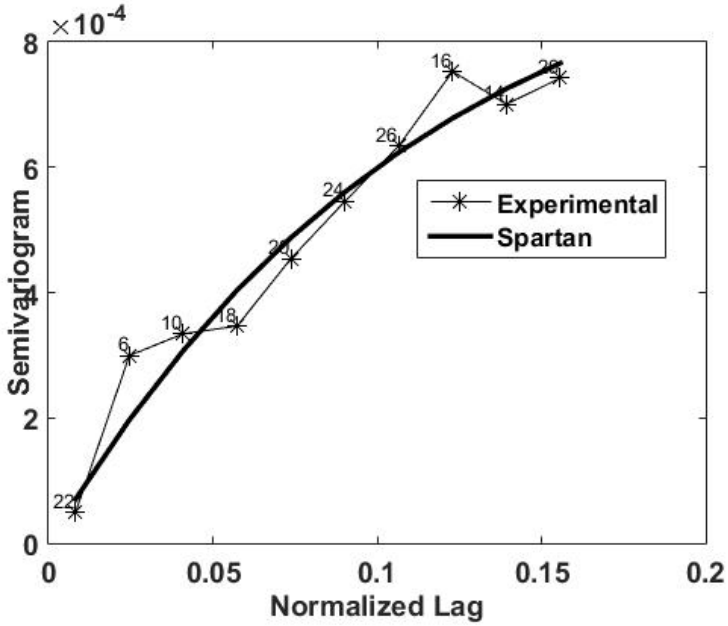
Lastly, the effect of the Akaike minimization to the corresponding variogram is demonstrated, and then, the corresponding mappings.



(a) Spartan akaike variogram 30

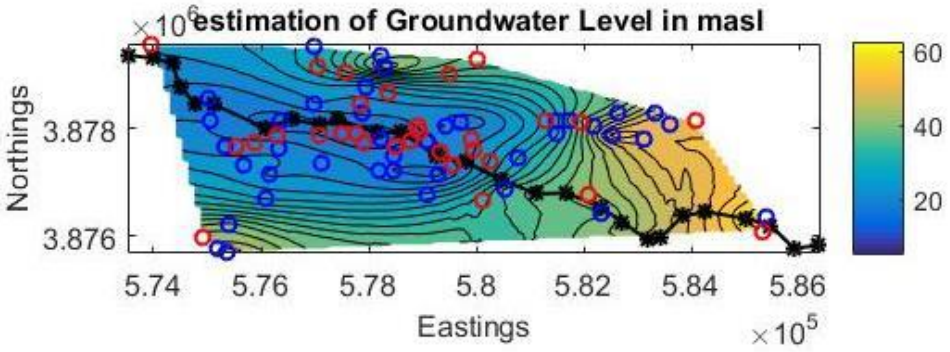


(b) spartan akaike variogram 40

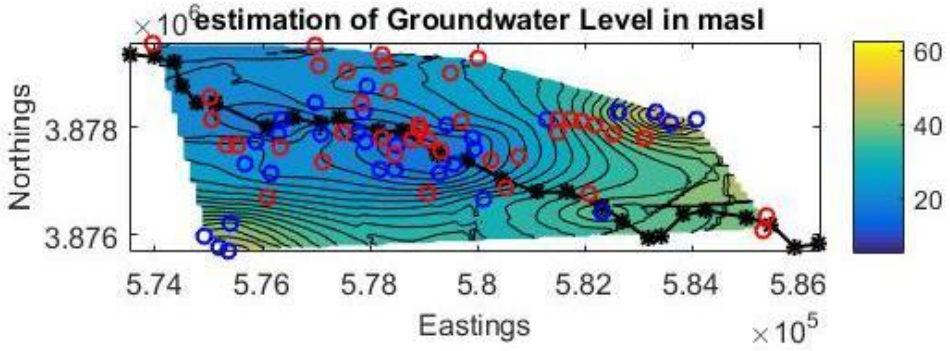


(c) spartan akaike variogram 50

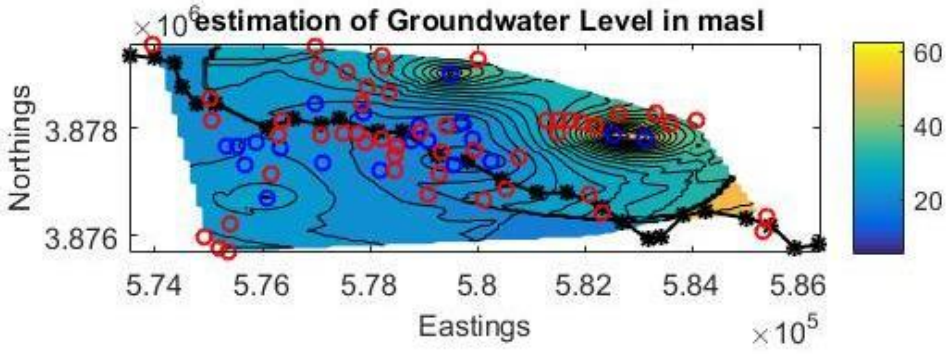
Figure 18: akaike minimization scenarios variogram



(a) Spartan akaike 30



(b) spartan akaike 40



(c) spartan akaike 50

Figure 19: Akaike minimization scenarios mapping

Here, the drawback of the Akaike minimization is that, it targets on the best candidate that has the best variogram fit, and as a result, the experimental variogram is more smooth, so that the smooth curve of the theoretical can interpolate better. So, in the corresponding mappings, there is an obvious simplicity and smoothness, because it is based on a less steep variogram.

5.2 Genetic Algorithm Stability-Sensitivity analysis (SP)

Other parameters that can be specified by experiment are the Stall generation termination criterion, and the population of each generation at an experiment. For that matter a sensitivity analysis for the spartan case is presented, in the following table. A viable scenario was selected for this, and that is the 30 removal scenario.

Table 8: Sensitivity analysis (30 removals) RMSD error spartan variogram

	RMSD	Gen	Time/Gen	Approx. Total Time
Pop=50,Stall=20	0.8690	97	100 sec	9700 sec
Pop=50,Stall=40	0.8145	98	100 sec	9800 sec
Pop=100,Stall=20	0.7618	60	226 sec	13560 sec
Pop=100,Stall=40	0.5785	98	226 sec	22148 sec
Pop=200,Stall=20	**0.5671**	72	410 sec	29520 sec
Pop=200,Stall=40	0.5543	86	410 sec	35220 sec
Pop=400,Stall=20	0.5445	67	830 sec	55610 sec

In theory, bigger stall generations would increase the chances of achieving a better result. But it is observed that after a certain threshold of pop=200, stall=20, a slightly better improvement is made, but with a tremendous increase of total time. Secondly , a similar table is shown for the RMSE which is very faster as mentioned because it does not need computing in the whole grid, just on the missing points. Furthermore, the number of evaluation of the fitness function (Fval) is being shown, which is an indicator of the convergence of the algorithm.

Table 9: Sensitivity analysis (30 removals) RMSE error spartan variogram

	RMSE	Fval	Total Time
Pop=50,Stall=20	11.0236	2100	76.1 sec
Pop=50,Stall=40	7.34644	10400	410.9 sec
Pop=50,Stall=60	10.9571	8350	310.1 sec
Pop=100,Stall=20	8.9570	6200	241.37 sec
Pop=100,Stall=40	8.5808	11000	427.05 sec
Pop=100,Stall=60	**6.6969**	21000	741.1 sec
Pop=200,Stall=20	**6.8550**	14200	509.2 sec
Pop=200,Stall=40	9.0513	14000	496 sec
Pop=200,Stall=60	7.1665	47600	1723.2 sec
Pop=400,Stall=20	7.8840	22000	811.8 sec
Pop=400,Stall=40	7.7831	35200	1351 sec
Pop=400,Stall=60	**6.7133**	90400	3171 sec

It is observed that in the experiments that needed very little time, the error is bigger than expected, and that is because the algorithm,because of either a little population or a small stall, did not get the chance to evolve to a smaller error, until the requirements of the termination criterion was met. Lastly, a similar sensitivity analysis for the Akaike criterion is being demonstrated .

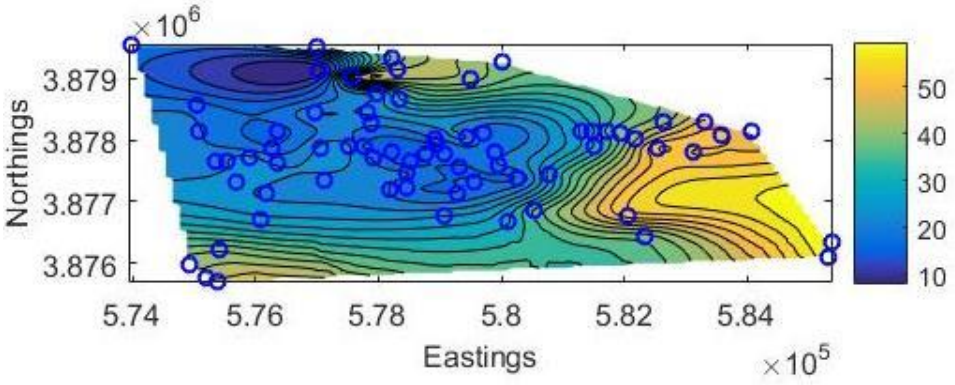
Table 10: Sensitivity analysis (30 removals) Akaike error spartan variogram

	Akaike	Fval	Total Time
Pop=100,Stall=20	-160.3098	3800	136.4 sec
Pop=100,Stall=40	-161.4695	7800	227.4 sec
Pop=100,Stall=60	**-173.9747**	16700	595.7 sec
Pop=200,Stall=20	-156.2765	10800	365.1 sec
Pop=200,Stall=40	-159.0287	14000	536.1 sec
Pop=200,Stall=60	**-169.9933**	23400	834.1 sec
Pop=400,Stall=20	-167.4112	18800	1457.6 sec
Pop=400,Stall=40	**-168.7725**	41600	1604.3 sec
Pop=400,Stall=60	-157.2270	30400	1046.6 sec

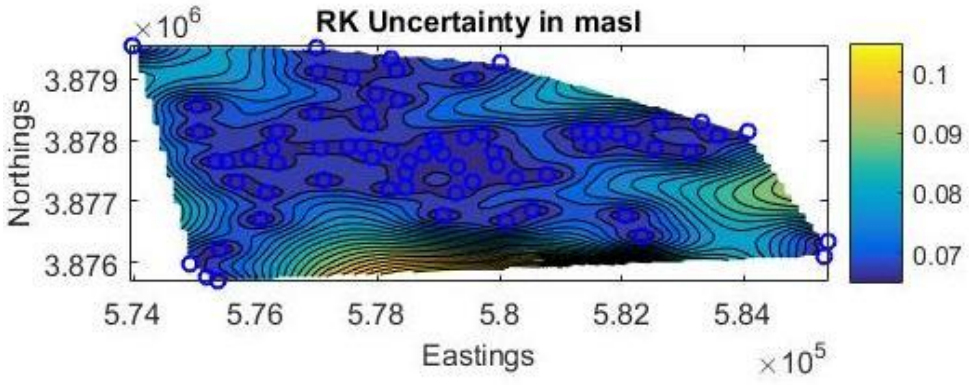
Here it seems to be an even faster optimization, because there is no need for neither cross validation (RMSE) or calculation of kriging in the entire domain (RMSD). The downside is that the mappings are not as good as with the other 2 errors (in comparison with the original that is), but with akaike, an almost perfect fit in some cases of the theoretical variogram to the experimental is being done, and so, the genetic algorithm chooses to keep the data that does not have many deviations with one another. As a result, a much more smoother transition at the reduced mapping is observed, which is, in some cases, not very realistic.

5.3 Powerlaw Optimization

In addition, similar results with the Powerlaw variogram are presented. Firstly, there are the initial mapping and corresponding kriging variance (uncertainty)

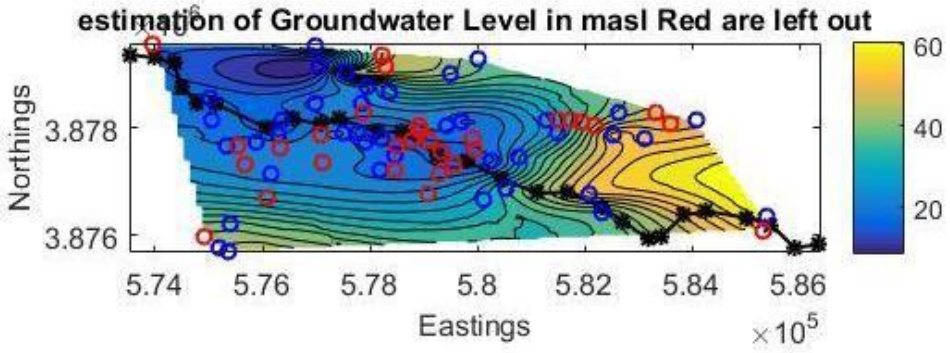


(a) Powerlaw initial mapping

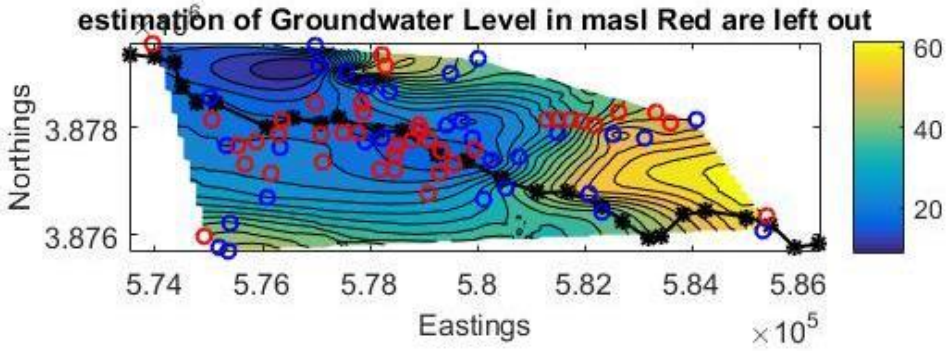


(b) Powerlaw initial uncertainty

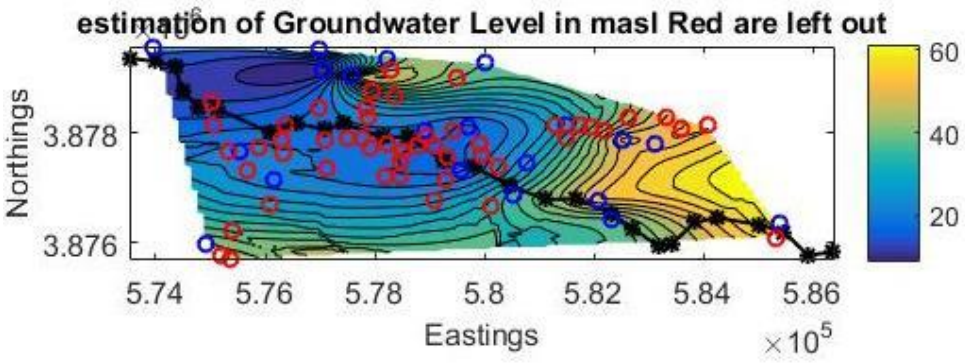
Figure 20: Powerlaw 70 data initial mapping and uncertainty



(a) Power RMSD 30

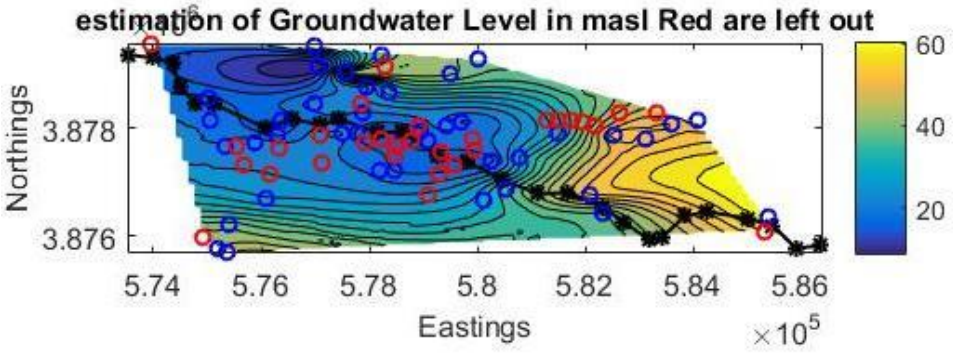


(b) Power RMSD 40

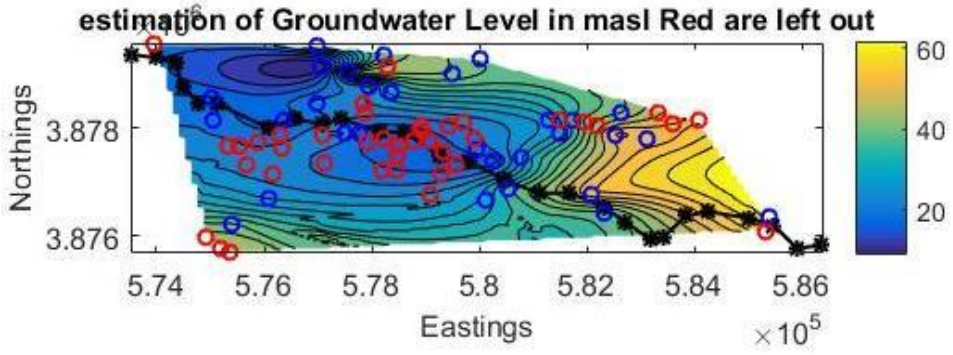


(c) Power RMSD 50

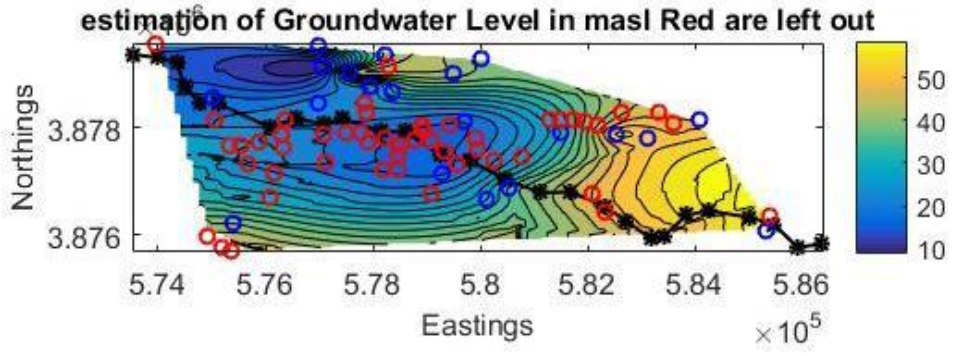
Figure 21: RMSD minimization scenarios mapping



(a) Power RMSE 30



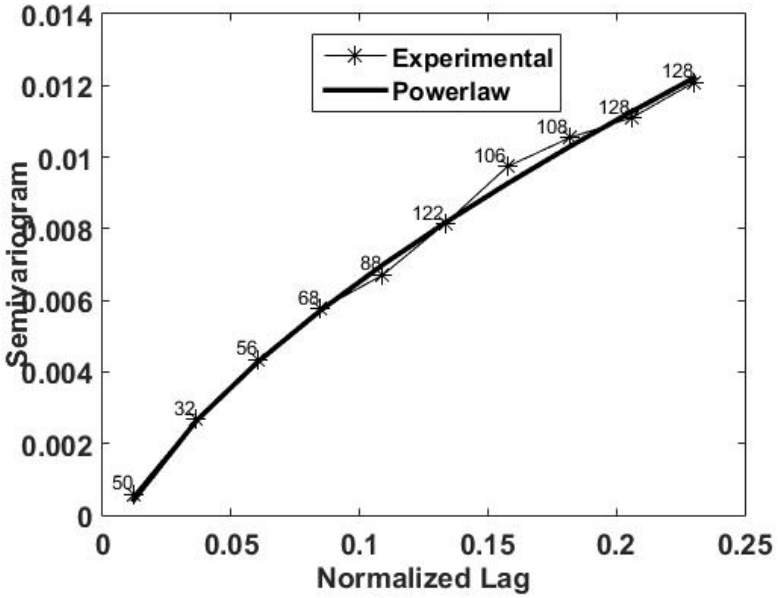
(b) Power RMSE 40



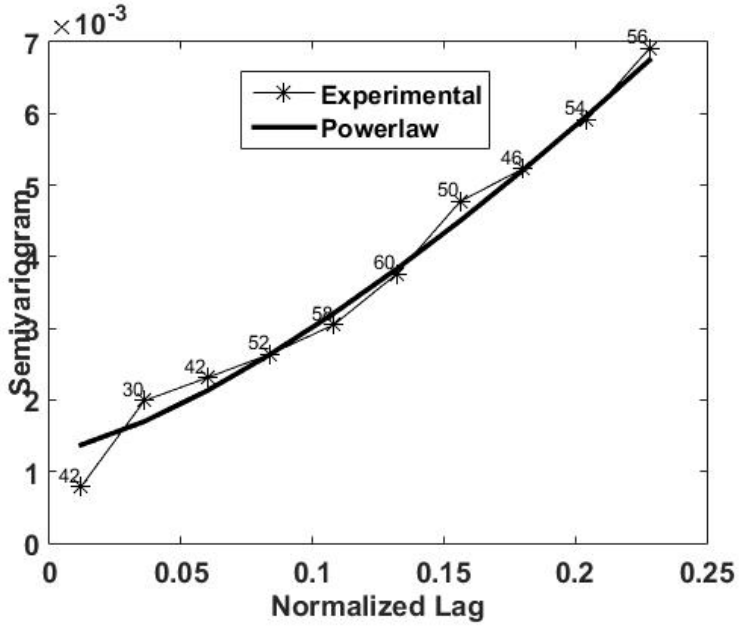
(c) Power RMSE 50

Figure 22: RMSE minimization scenarios mapping

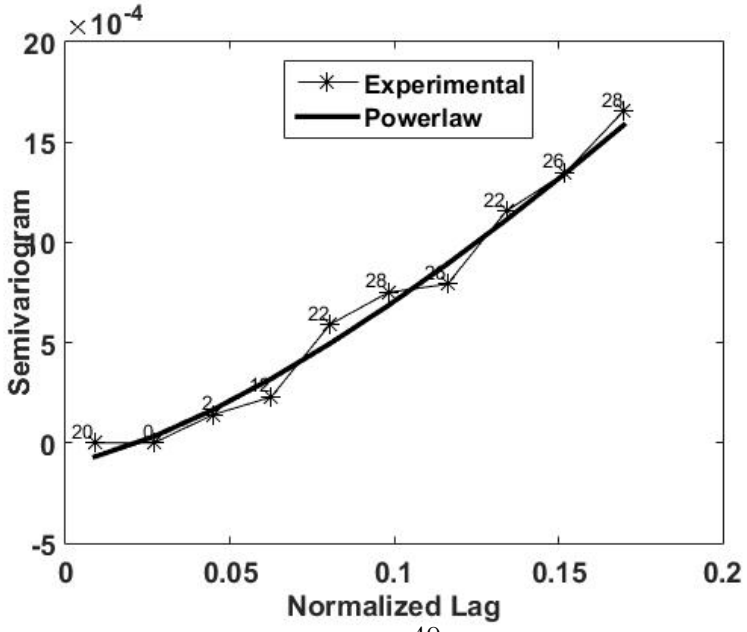
In RMSD and RMSE minimizations, there is a very good resemblance of the initial mapping, as indicated by Table 7, but with a little more error than Spartan optimization.



(a) Power Akaike variogram 30



(b) Power Akaike variogram 40

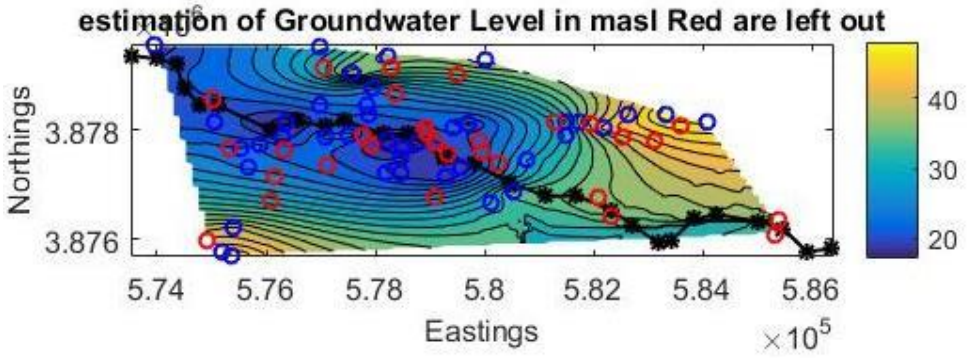


(c) Power Akaike variogram 50

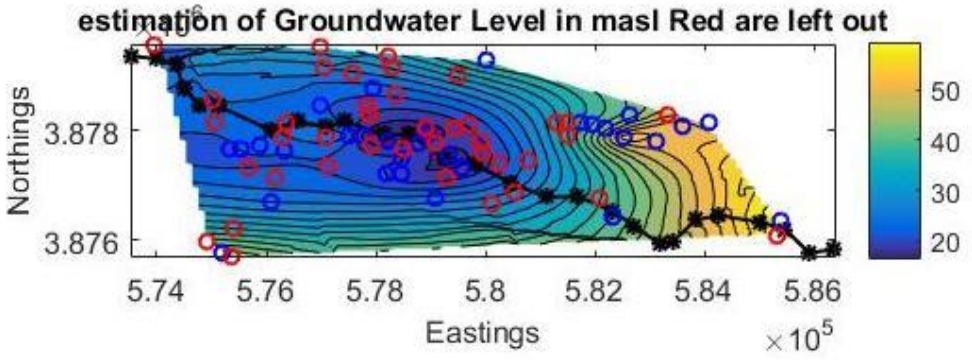
Figure 23: Akaike minimization scenarios variogram

In the above figures, as expected from the Akaike minimization, the chosen candidate to remove is that, the experimental variogram is easily fitted by the theoretical Powerlaw variogram.

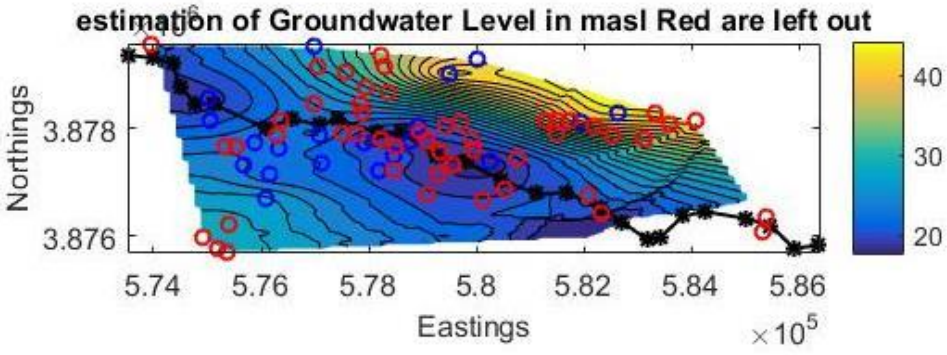
In the next figures, the corresponding mappings from the 3 above variograms are presented. A choice that tends to remove measurements from the center that is more monitored is observed, as well as a tendency to smoothen the irregularities.



(a) Power akaike 30



(b) Power akaike 40



(c) Power akaike 50

Figure 24: Akaike minimization scenarios mapping

6 Test case 2: Drama

The similar test case of Drama is being presented at this section, with the difference that the initial borehole measurements of underwater level are 250 and that these measurements have a lower range than Mires. As presented earlier, an analogous table with the errors which are minimized is being made, and the deviation from the initial mapping, when RMSD is not optimized genetically.

Table 11: SPARTAN OPTIMIZATION

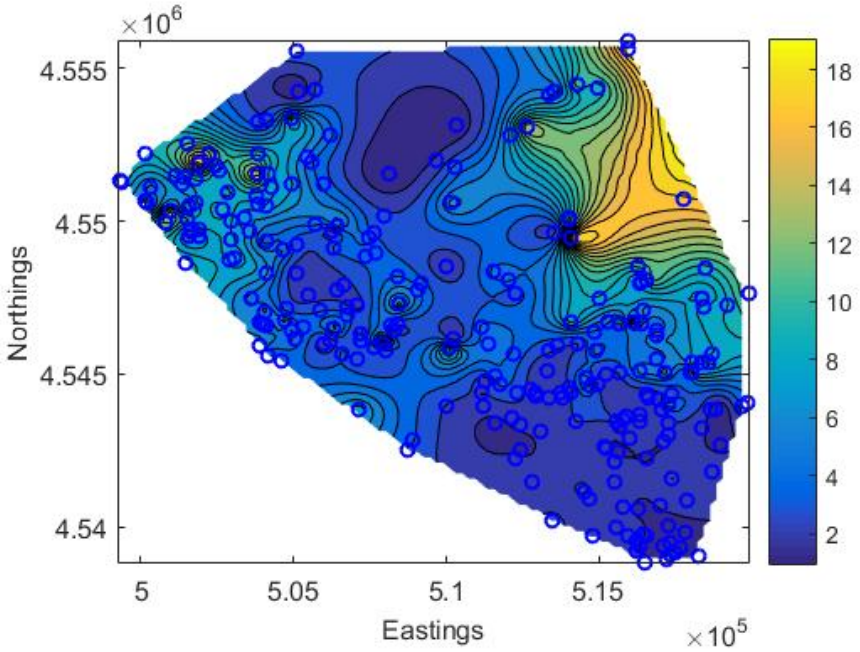
Scenarios :	150	200	220
RMSD optimization	0.4160	0.6980	0.8145
RMSE optimization	11.6747	25.4796	32.0131
Corresponding RMSDs	0.3815	0.8173	0.8697
Akaike optimization	-157.27	-137.89	-161.5690
Corresponding RMSDs	1.5957	1.9099	3.1063

Table 12: POWERLAW OPTIMIZATION

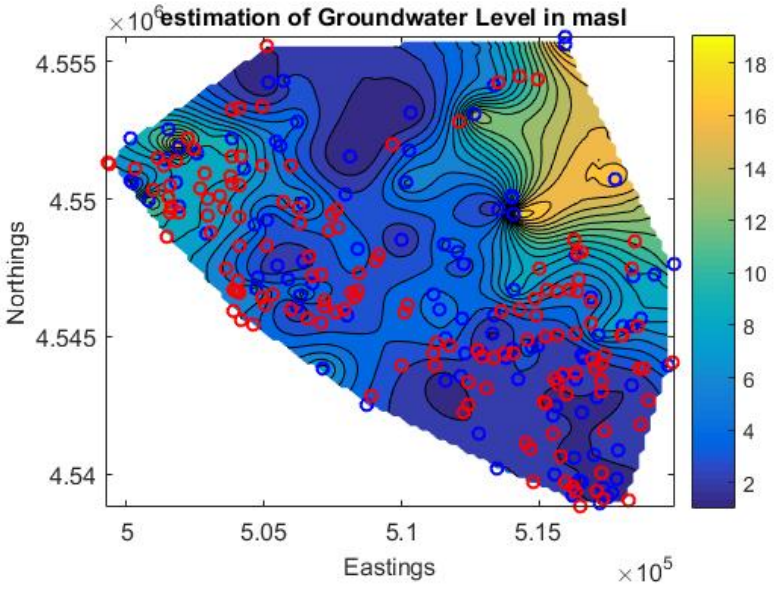
Scenarios :	150	200	220
RMSD optimization	0.4176	0.9046	1.0442
RMSE optimization	13.3035	22.8317	29.2525
Corresponding RMSDs	1.6055	1.6196	1.7216
Akaike optimization	-183.02	-165.85	-164.53
Corresponding RMSDs	1.7916	2.4325	2.8975

The first figures are the initial mapping with Spartan variogram and the following figures are the results from RMSD, RMSE and Akaike optimization, subtracting 150,200 and 220 data points from our 250 initial data.

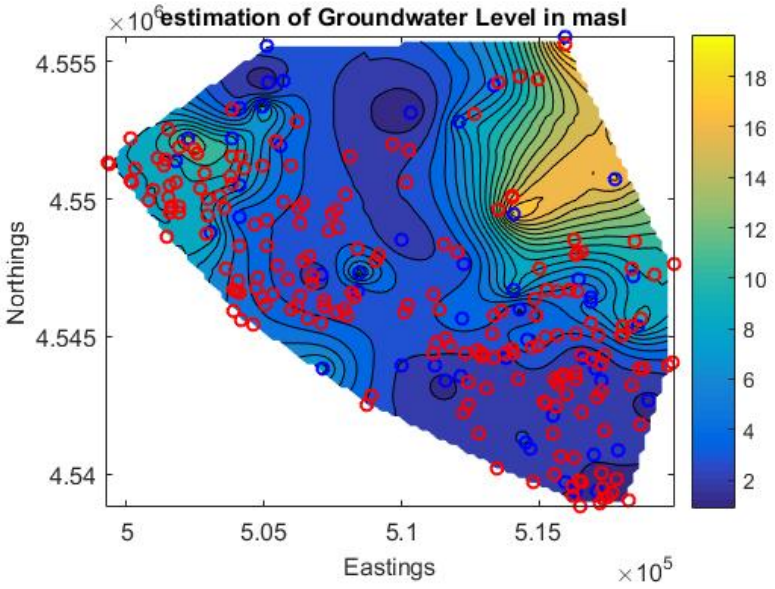
6.1 Spartan variogram



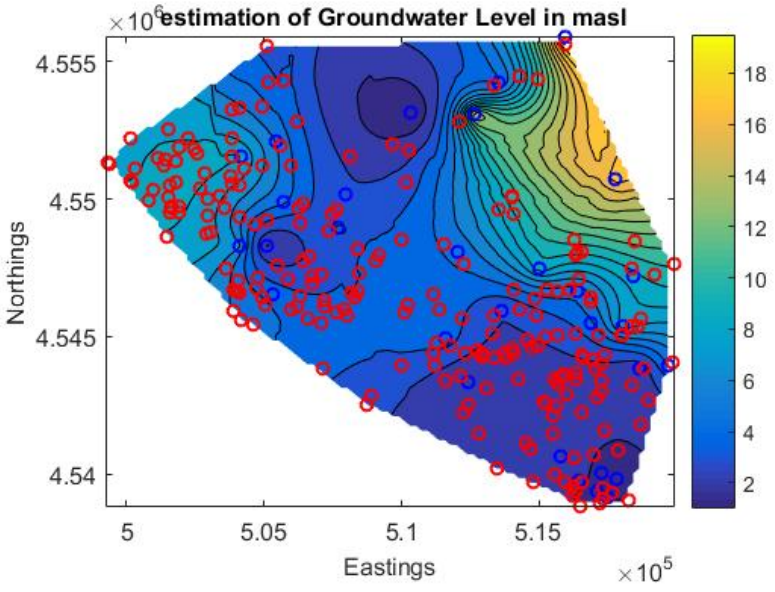
(a) spartan initial 250 measurements
Figure 25: Initial 250 data mapping



(a) Spartan RMSD 150

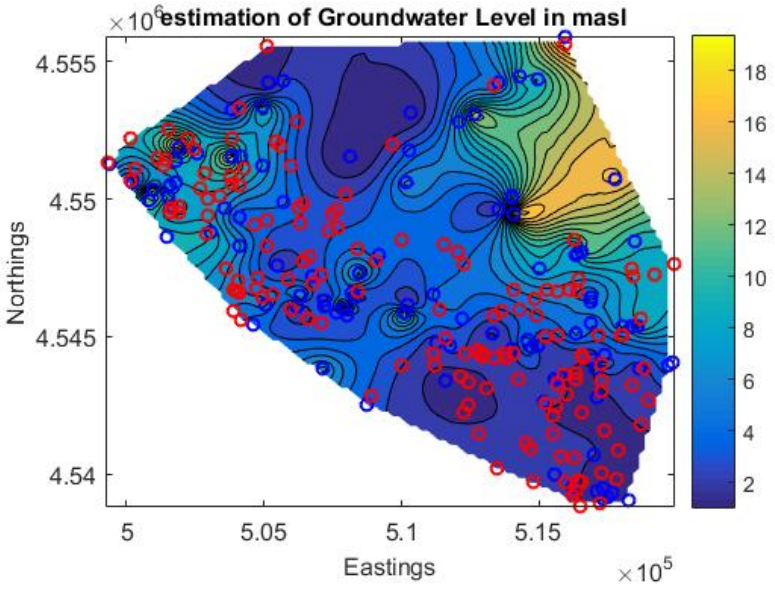


(b) spartan RMSD 200

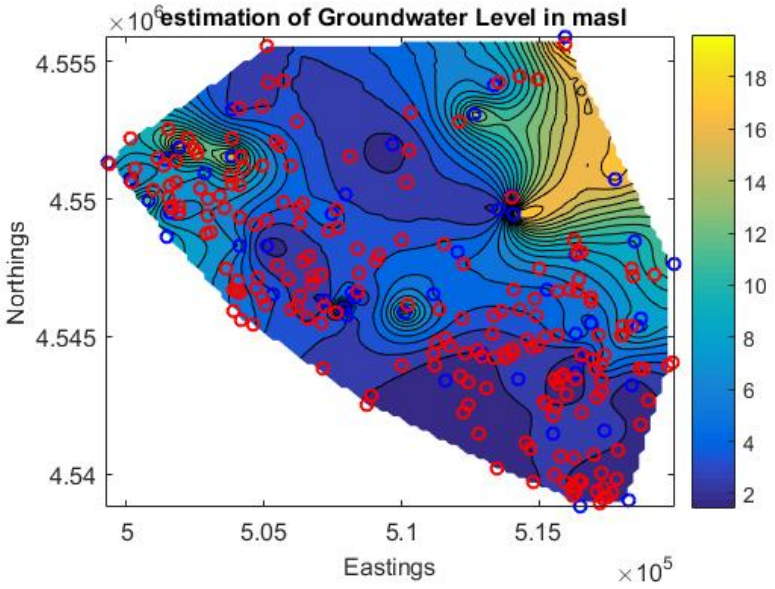


(c) spartan RMSD 220

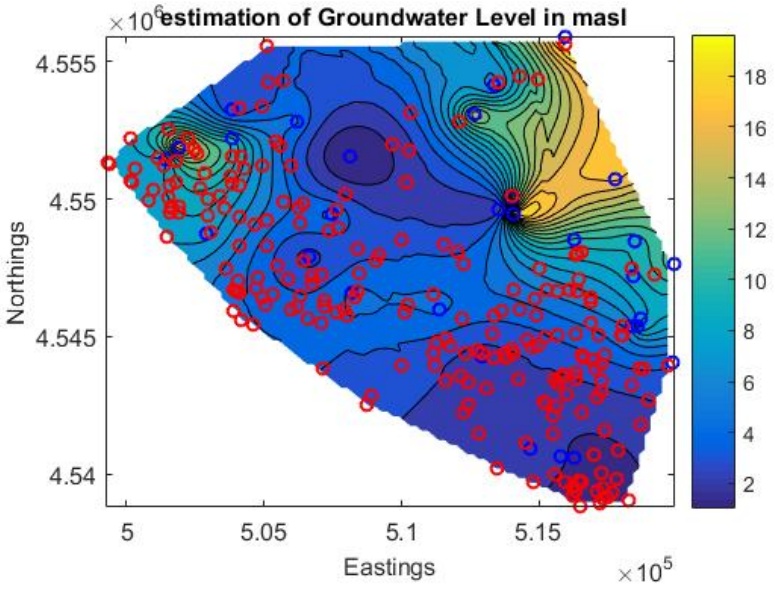
Figure 26: RMSD minimization scenarios mapping



(a) Spartan RMSE 150



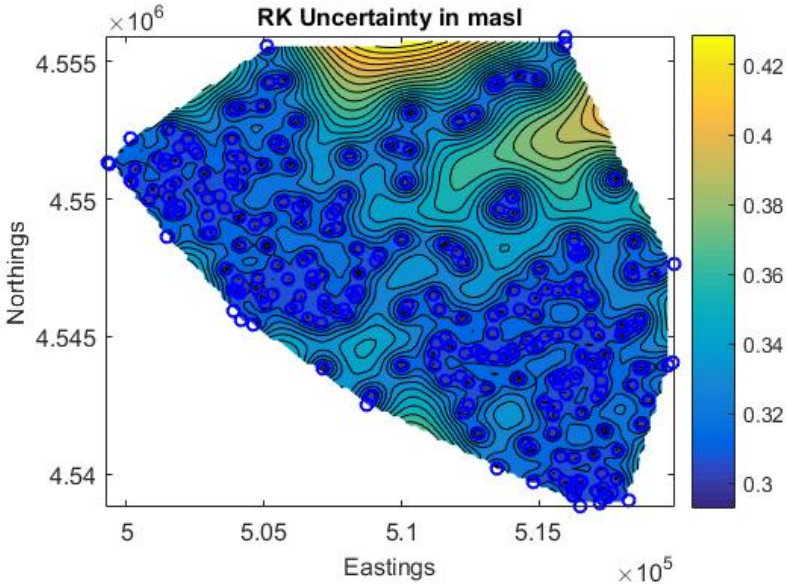
(b) spartan RMSE 200



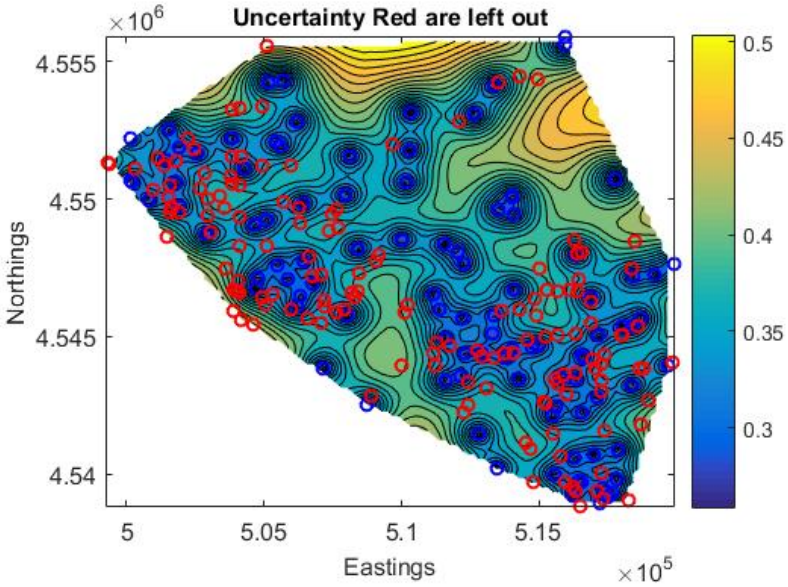
(c) spartan RMSE 220

Figure 27: RMSE minimization scenarios mapping

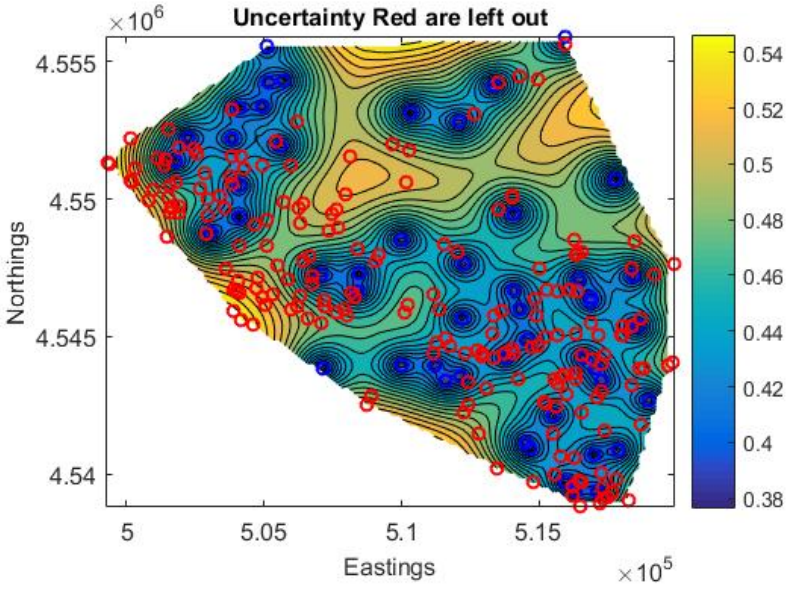
In the above figures, the scenarios of 150,200 and 220 removals are shown out of the 250 initial measurements, with the RMSE and RMSD minimizations, which, like the first test case, resembles the original mapping. In the 220 case, more and more differences can be observed, (as indicated by table 11) and as a result, the proposed scenario is to keep at least 50 out of the 250 measurements (which is still impresive keeping only 20 % of the initial mapping measurements and still have an adequate mapping.)



(a) spartan initial uncertainty



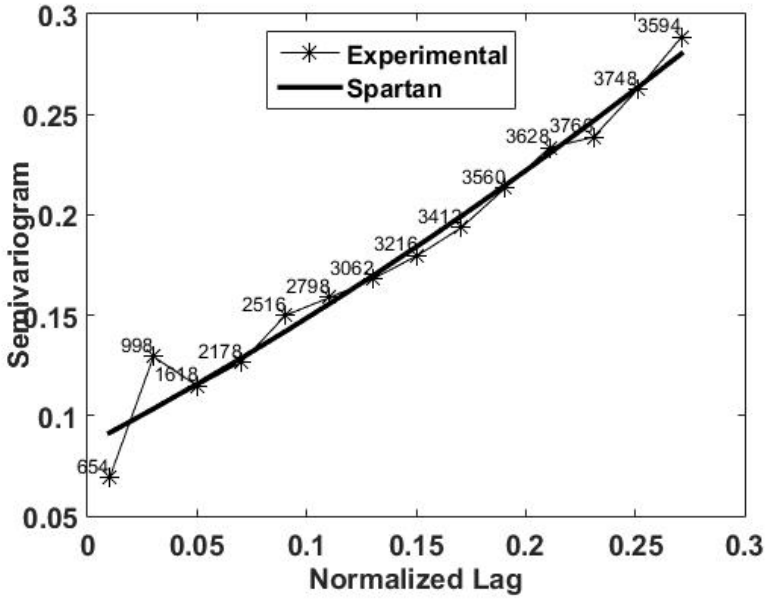
(b) spartan RMSD uncertainty 150



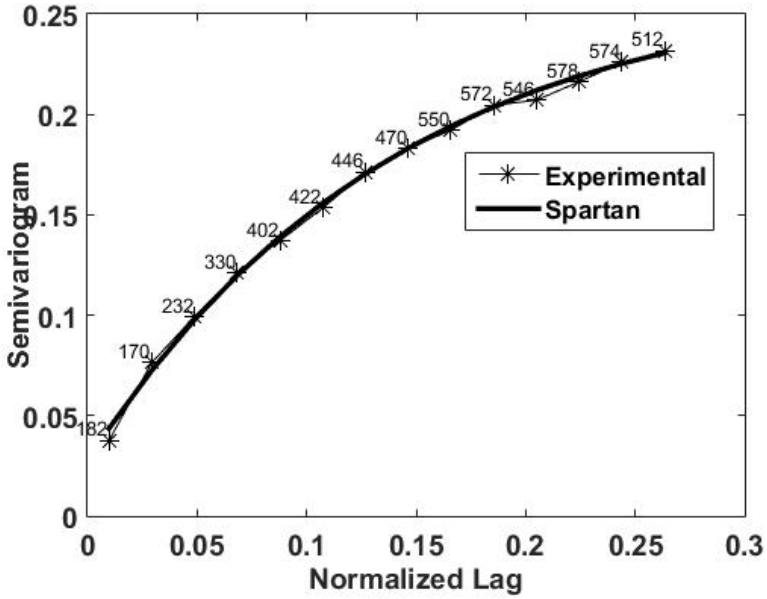
(c) spartan RMSD uncertainty 200

In the above figure it is observed what we may have expected, that the more data points are removed, the more the uncertainty grows, and the mapping shows where one would need more sampling for the inverse purpose of this project.

Lastly, the effect of the Akaike minimization is observed, to the corresponding variograms along with the original 250 data spartan variogram, and then, the corresponding mappings.

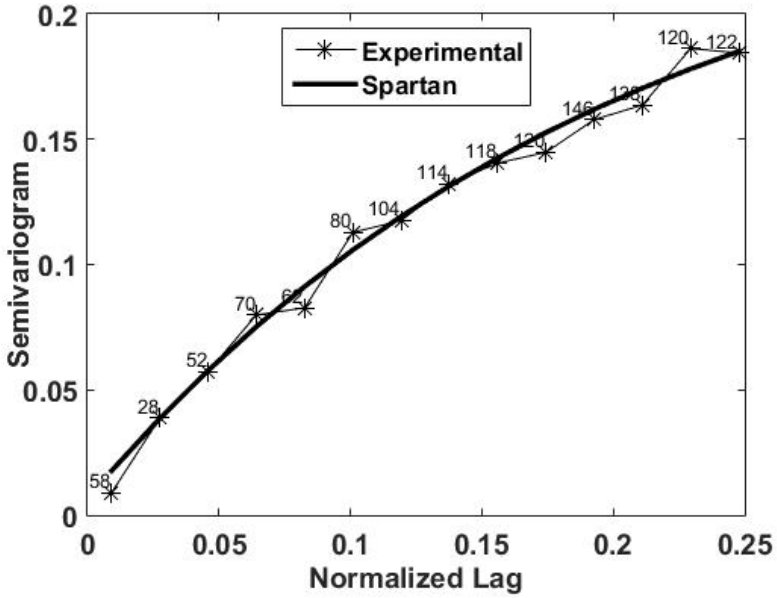


(a) Spartan initial variogram

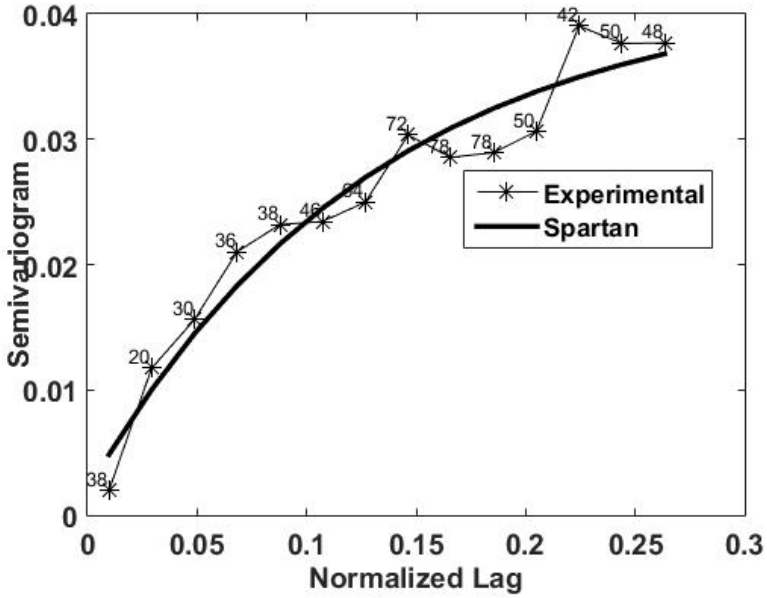


(b) Spartan variogram akaike optimization 150 removals

Figure 29: akaike minimization initial and scenarios variogram

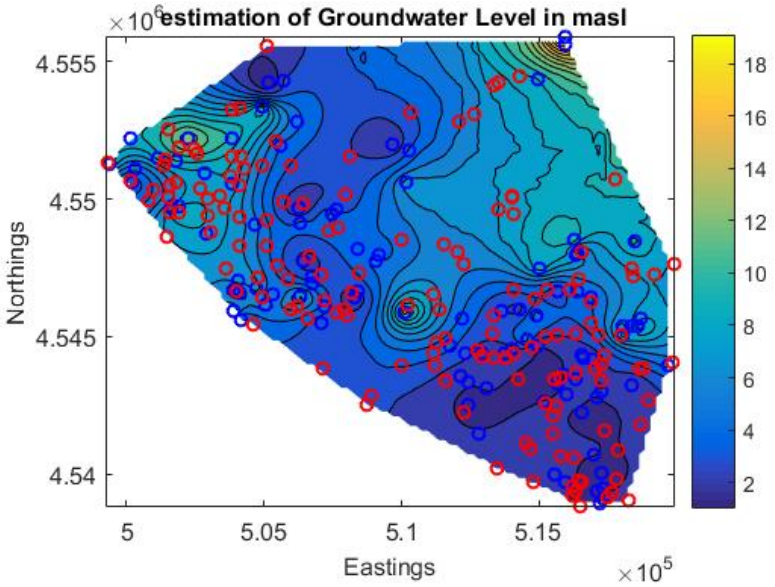


(a) Spartan variogram akaike optimization 200 removals

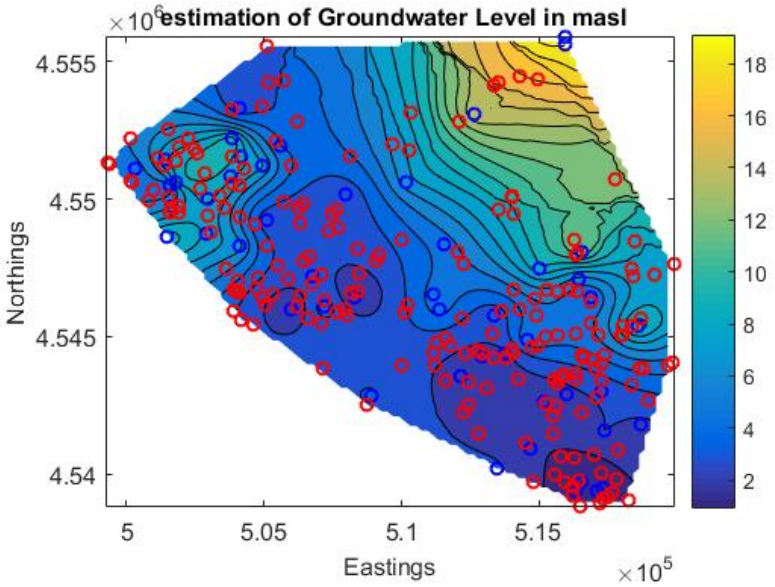


(b) Spartan variogram akaike optimization 220 removals

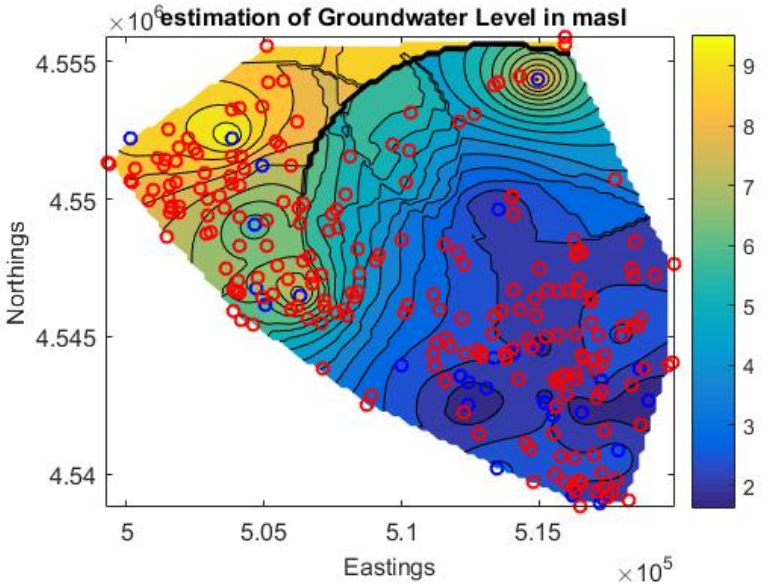
In the above figures, not only is observed the perfect fit on the 150 and 200 removals, but also, the flexibility of the Spartan variogram is shown clearly, because, depending on the rigidity coefficient, depending on the needs that the data indicate, it can be convex or concave, or have a sinusoidal form to fit more complex data. As a result, Spartan can be utilized more easily in many test cases and different applications.



(a) Spartan akaike 150



(b) spartan akaike 200

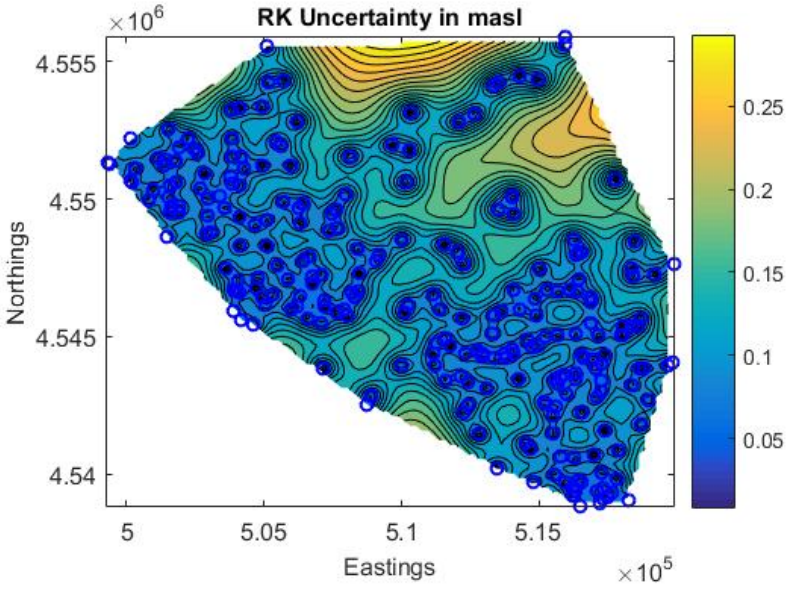


(c) spartan akaike 220

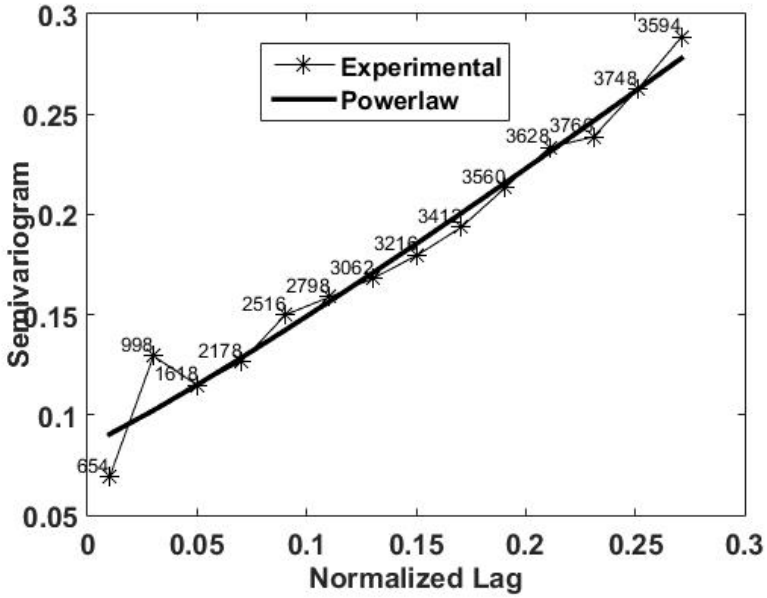
Figure 31: akaike minimization scenarios mapping

6.2 Powerlaw Optimization

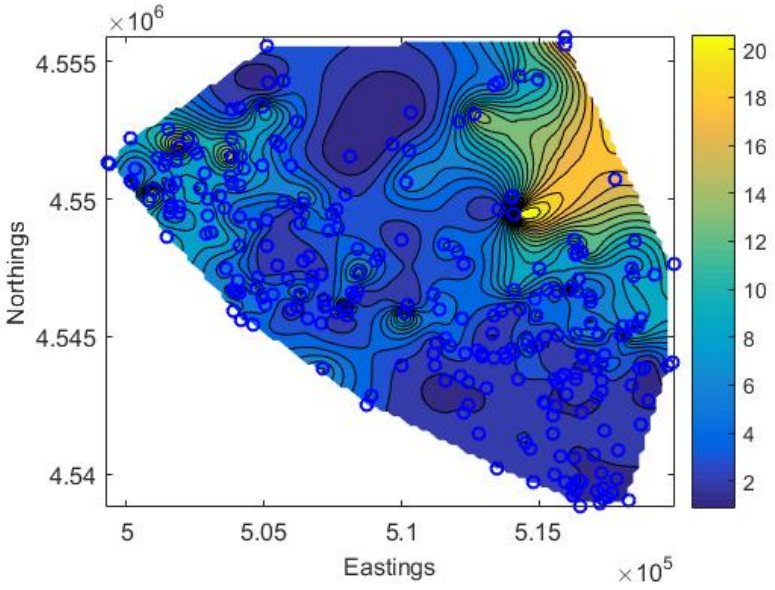
The powerlaw initial uncertainty, variogram and mapping alongside with the results from our 3 measures follows.



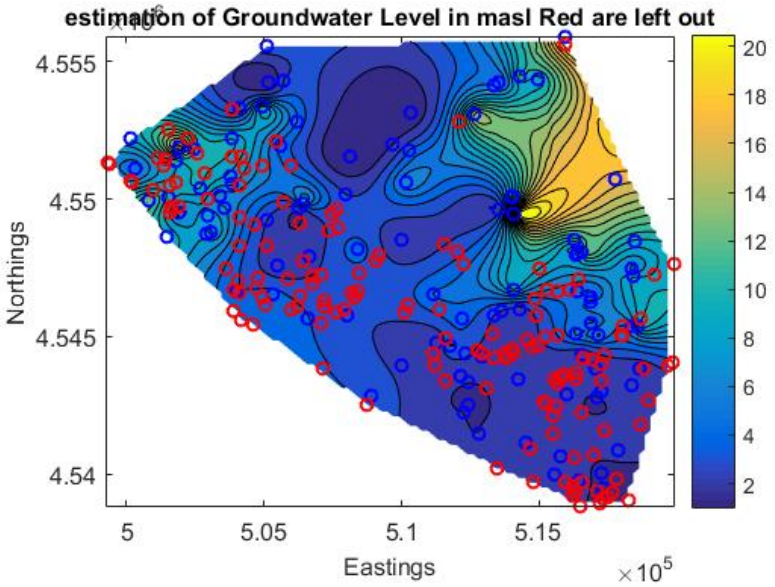
(a) power initial uncertainty



(b) power initial variogram

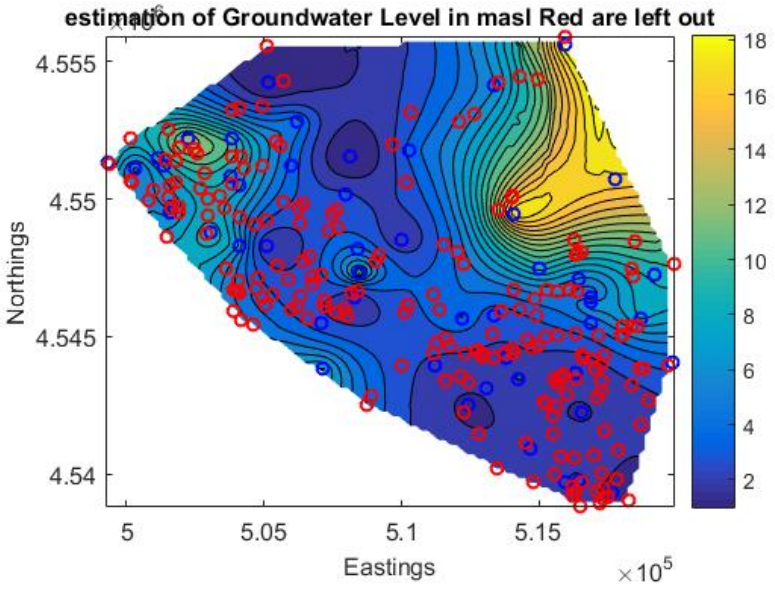


(a) power initial 250 data mapping

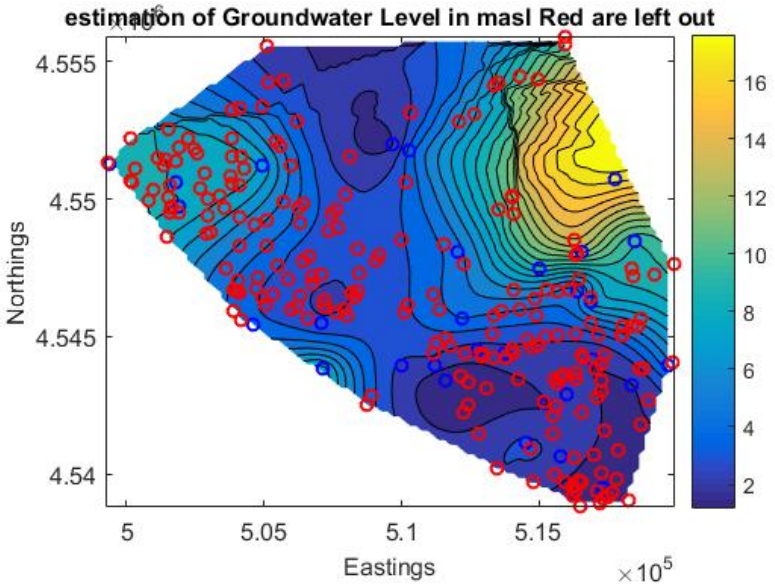


(b) power RMSD 150

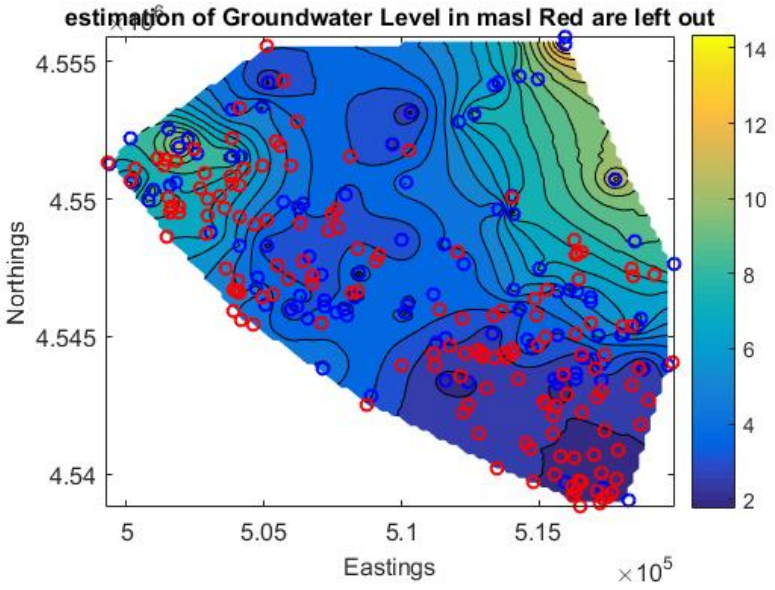
Figure 33: Initial mapping and RMSD optimized scenarios mapping



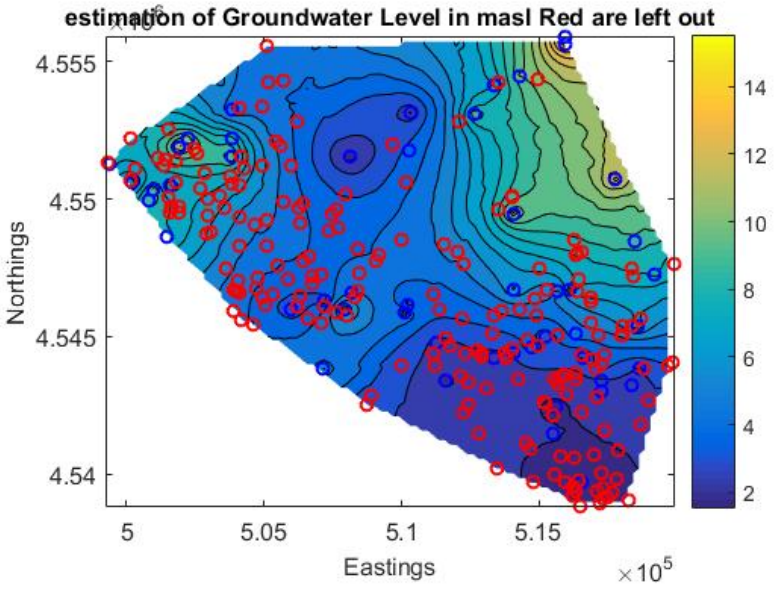
(a) power RMSD 200



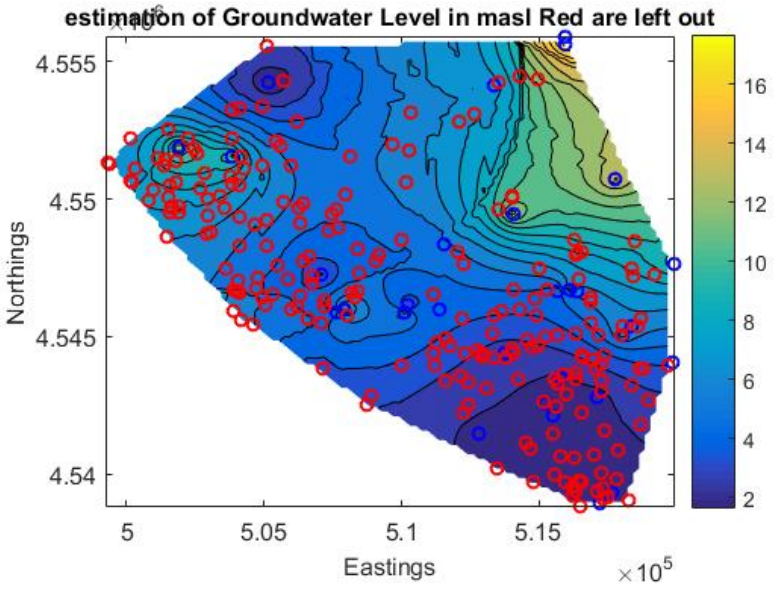
(b) power RMSD 220



(a) power RMSE 150



(b) power RMSE 200

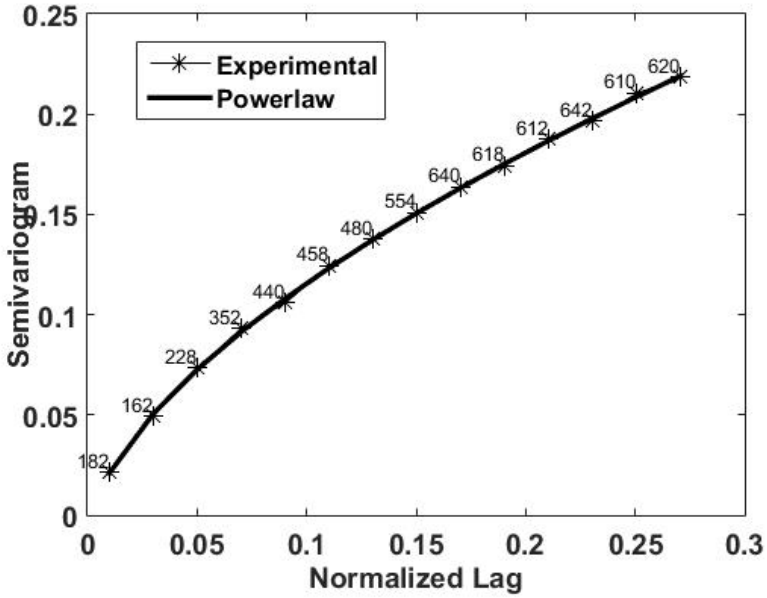


(c) power RMSE 220

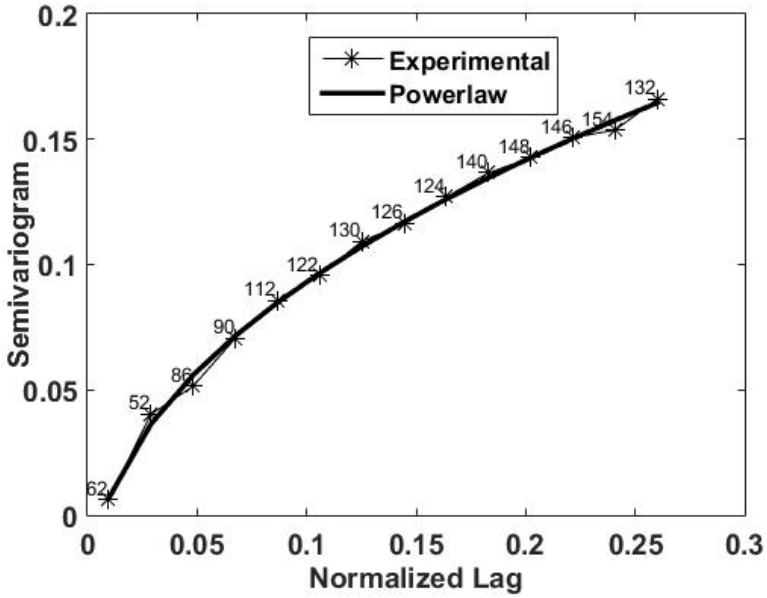
Figure 35: RMSE minimization scenarios mapping

In the RMSD and RMSE variograms a worse situation is being observed in comparison with the Spartan case, as indicated by the comparison of tables 11 and 12, that shows RMSD optimization that finds better minimum with the spartan optimization, and also in the RMSE and Akaike optimization, the corresponding RMSDs (Mean Deviation of the initial map from the reduced one), the Spartan outperforms the Powerlaw in almost every instance.

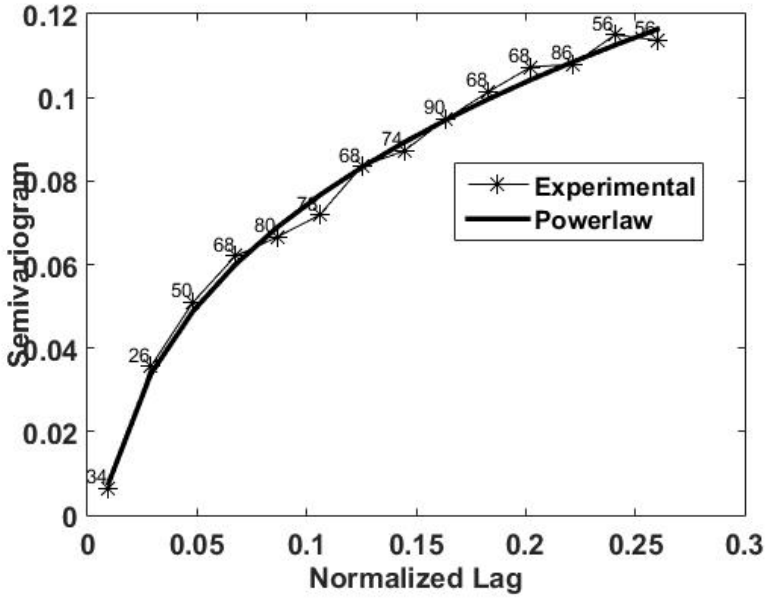
In the figures below, the astounding fitting of the Powerlaw Akaike minimization is observed on the variogram, but that does not translates to a very good mapping in comparison with the initial, as expected, especially in the 220 scenario where the mapping is misleading, so it may be proposed to avoid so much under sampling.



(a) power akaike variogram 150

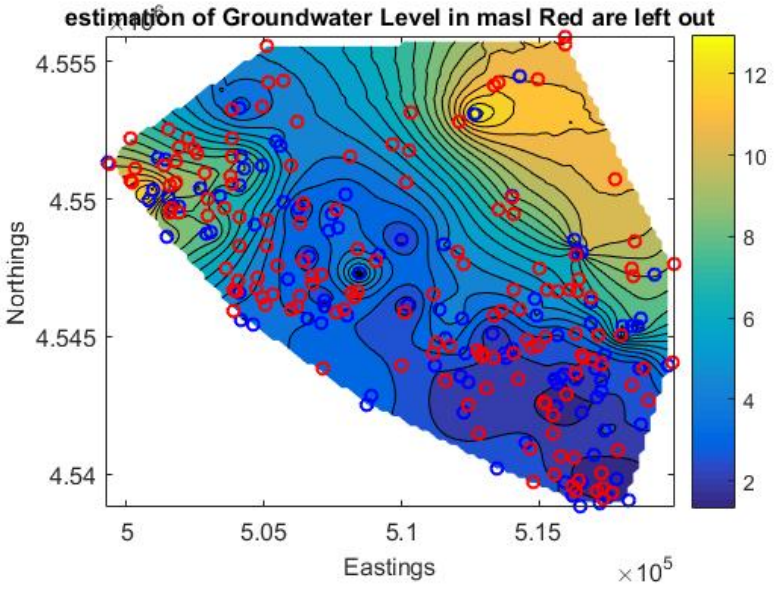


(b) power akaike variogram 200

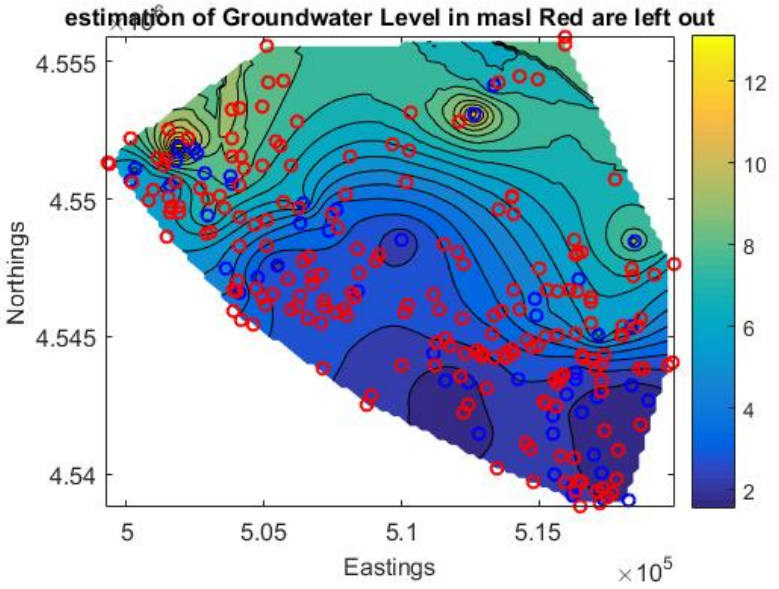


(c) power akaike variogram 220

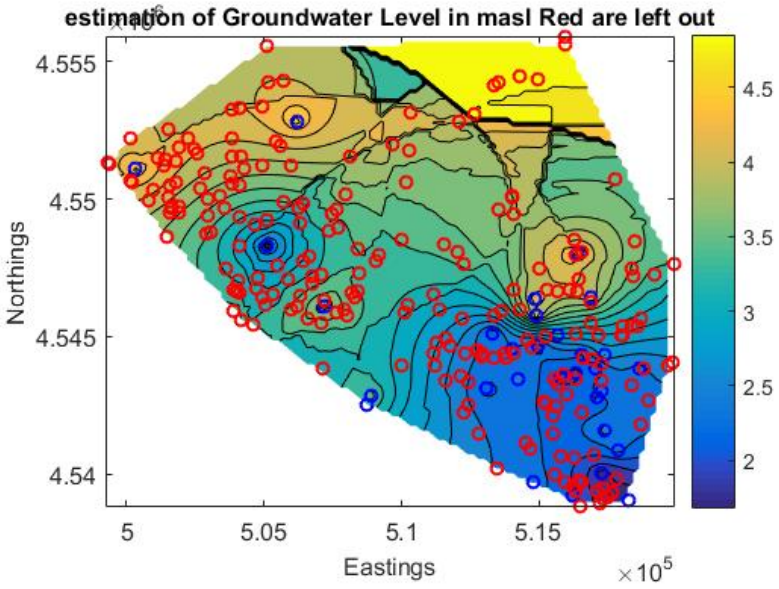
Figure 36: akaike minimization scenarios variogram



(a) power akaike 150



(b) spartan akaike 200



(c) power akaike 220

Figure 37: akaike minimization scenarios mapping

7 Future Work-Discussion

In conclusion, in this work, two test cases of monitored groundwater networks are presented, in which our basins were heavily monitored, and it was possible to subtract more than 50% of Mires basin measurements and 70-80% of Dramas measurements (In Drama the water level had very little range in comparison to Mires and the measurements were closer to one another.) Possible future work that can be made in this coupled methodologies tool, is to make the evaluation work in parallel, either in the GPU, or in Matlab's parallel workers, so that even RMSD can be fast and efficient in practice. Another improvement could be to use multi objective genetic algorithms so that one could minimize with respect to everyone of the errors and choose from a Pareto front the minimizer that suits the needs. A simpler multi objective optimization can be done by defining the fitness function as in [15], by a weighted linear combination of the errors defined, in which the weights are chosen experimentally according to the needs of the minimization. There, kriging variance could be added as one of the errors, minimizing without the box cox transform, because when the normalization is being made, an exponent is sought, and then our data can be more evenly distributed, and using that exponent, we can back transform the predicted level at the unsampled locations. But one can not back transform the kriging variance as it has no units. So there is no reason for this work to minimize in respect to kriging variance, because of the single objective and the transformation of the data that is presented here.

8 Bibliography

[1] Aboufirassi M, Marino M. Kriging of water levels in the Souss aquifer, Morocco. Math Geol 1983;15:537–51.

[2] Ahmadi S, Sedghamiz A. Geostatistical analysis of spatial and temporal variations of groundwater level. Environ Monit Assess 2007;129:277–94.

[3] Beckley, M.C., Comparison of Sampling Methods for Kriging Models, Master Thesis- University of Pretoria, 2014

[4] Box GEP, Cox DR. An analysis of transformations. J R Stat Soc Ser B 1964;26: 211–52.

[5] Buchanan, S., Triantafilis, J. (2009). Mapping water table depth using geophysical and environmental variables. Ground Water, 47(1), 80–96.

[6] Cameron, Kirk, Hunter, Philip, 2002, Using spatial models and kriging techniques to optimize long-term ground-water monitoring networks—A case study: Environmetrics, v. 13, p. 629–656.

[7] Chander S, Kumar A, Spolia SK. Flood frequency analysis by power transformation. J Hydraul Div: Proc ASCE 1978;104:1495–504.

[8] Cressie N., The Origins of Kriging, Mathematical Geology, Vol. 22, No. 3, 1990

[9] Delhomme JP. La cartographie d'une grandeur physique a partir des donnees de differentes qualities. In: International Association of Hydrogeologists, editor. Proceedings of IAH congress. Montpelier, France: IAH; 1974. p. 185–94.

[10] Delhomme JP. Kriging in the hydrosciences. Adv Water Resour 1978;1:251–66.

[11] Desbarats AJ, Logan CE, Hinton MJ, Sharpe DR. On the kriging of water table elevations using collateral information from a digital elevation model. J Hydrol 2002;255:25–38.

[12] Deutsch CV, Journel AG. GSLIB. Geostatistical software library and user's guide. New York: Oxford University Press; 1992.

[13] Dhar, Anirban, and Datta, Bithin, 2010, Logic-based design of groundwater monitoring network for redundancy reduction: Journal of Water Resources Planning and Management, v. 136, p. 88–94.

[14] Donta, A.A., Lange, M.A., and Herrmann, A., 2006. Water on Mediterranean islands: Current conditions and prospects for sustainable management. Project No EVK1-CT-2001-00092-Funded by the European Commission, ISBN 3-9808840-7-4. Münster: Centre for Environment Research (CER), University of Münster.

[15] Fisher, J. C., Optimization of Water-Level Monitoring Networks in the Eastern Snake River Plain Aquifer Using a Kriging-Based Genetic Algorithm Method, Scientific Investigations Report 2013–5120, DOE/ID-22224

[16] Gambolati G, Volpi G. Groundwater contour mapping in Venice by stochastic interpolators 1. Theory Water Resour Res 1979;15:281–90.

[17] Gambolati G, Volpi G. A conceptual deterministic analysis of the kriging technique in hydrology. Water Resour Res 1979;15:625–9

[18] Goovaerts P. Geostatistics for natural resources evaluation. New York: Oxford University Press; 1997.

[19] Guan Jiabao, Elcin Kentel,Mustafa M. Aral, Genetic Algorithm for Constrained Optimization Models and Its Application in Groundwater Resources Management,Journal of Water Resources Planning and Management, Vol. 134, No. 1, January 1, 2008. ©ASCE, ISSN 0733-9496/2008/ 1-64–72

[20] Hessami,M.,Anctil,F., Viau,A.A.(2001).Delaunay implementation to improve kriging computing efficiency. Computers and Geosciences, 27(2), 237–240.

[21] Hirsch RM. Synthetic hydrology and water supply reliability. Water Resour Res 1979;15:1603–15.

[22] Hristopulos, D. T. (2003). Spartan Gibbs random field models for geostatistical applications. SIAM Journal on Scientific Computing, 24(6), 2125–2162.

[23] Hristopulos, D. T., Elogne, S. N. (2007). Analytic properties and covariance functions for a new class of generalized Gibbs random fields. IEEE Transactions on Information Theory, 53(12), 4667–4679.

[24] Hristopulos, D. T., Elogne, S. N. (2009). Computationally efficient spatial interpolators based on Spartan spatial random fields. IEEE Transactions on Signal Processing, 57(9), 3475–3487.

[25] Hoeksema RJ, Clapp RB, Thomas AL, Hunley AE, Farrow ND, Dearstone KC. Cokriging model for estimation of water table elevation. Water Resour Res 1989;25:429–38.

[26] Isaaks, E.H., and Srivastava, R.M., 1989, An introduction to applied geostatistics: New York, Oxford University Press, 561 p.

[27] Jain D, Singh VP. A comparison of transformation methods for flood frequency analysis. Water Resour Bull 1986;22:903–12.

[28] Kitanidis, P.K., 1997, Introduction to geostatistics-Applications to hydrogeology: New York, Cambridge University Press, 249 p

[29] Kumar V. Optimal contour mapping of groundwater levels using universal kriging-a case study. Hydrol Sci J 2007;52:1038–50.

[30] Matérn B. Spatial variation. Medd Fran Stat Skogsf-Institut 1960;49:1–144.

[31] Nunes, L.M., Cunha, M.C., and Ribeiro, L., 2004a, Groundwater monitoring network optimization with redundancy reduction: Journal of Water Resources Planning and Management, v. 130, p. 33–43

[32] Olea R, Davis JC. Optimizing the high plains aquifer water-level observation network. Open File Report 1999. Kansas Geological Survey, 1999.

[33] Pardo-Iguzquiza E, Chica-Olmo M. Geostatistics with the Matern semivariogram model: A library of computer programs for inference, kriging and simulation. Comput Geosci 2008;34:1073–9.

[34] Philip, G. M., Watson, D. F. (1986). Automatic interpolation methods for mapping piezometric surfaces. Automatica, 22 (6), 753–756.

[35] Prakash MR, Singh VS. Network design for groundwater monitoring-a case study. Environ Geol 2000;39:628–32.

[36] Pucci AAJ, Murashige JAE. Applications of universal kriging to an aquifer study in New Jersey. Ground Water 1987;25:672–8.

[37] Rouhani, S. (1986). Comparative study of ground-water mapping techniques. Ground Water, 24(2), 207–216.

[38] Salas J. Analysis and modeling of hydrologic time series. In: Maidment D, editor. Handbook of hydrology. New York, USA: McGraw-Hill; 1993. p. 19.11–72.

[39] Sophocleous M, Paschetto JE, Olea RA. Ground-water network design for northwest Kansas, using the theory of regionalized variables. Ground Water 1982;20:48–58.

[40] Sivananda, S.N., Deepa, S.N., Introduction to Genetic Algorithms, Springer 2008 ISBN 978-3- 540-73189- 4

[41] Srinivas, M., Patnaik, L.M., Adaptive Probabilities of Crossover and Mutation in Genetic Algorithms, IEEE transactions on systems, man and cybernetics, VOL. 24, NO. 4, PG 656, APRIL 1994

[42] Stein ML. Interpolation of spatial data. Some theory for kriging. New York: Springer; 1999.

[43] Sun, Y., Kang, S., Li, F., Zhang, L. (2009). Comparison of interpolation methods for depth to groundwater and its temporal and spatial variations in the Minqin oasis of northwest China. Environmental Modelling and Software, 24(10), 1163–1170.

[44] Theodoridou, P.G.,Varouchakis,E.A., Karatzas,G.P., Corzo Perez, G.A., Groundwater level geostatistical analysis using non-Euclidean distance metrics and variable variogram fitting criteria, 12th International Conference on Hydroinformatics, HIC 2016

[45] Theodossiou, N., Latinopoulos, P., 2006, Evaluation and optimisation of groundwater observation networks using the Kriging methodology: Environmental Modeling and Software, v. 21, no. 7, p. 991–1000

[46] Thyer M, Kuczera G, Wang QJ. Quantifying parameter uncertainty in stochastic models using the Box-Cox transformation. J Hydrol 2002;265: 246–57.

[47] Varouchakis, E. A., Integrated Water Resources Analysis at Basin Scale: A Case study in Greece, DOI:10.1061/(ASCE)IR.1943-4774.0000966, 2015 American Society of Civil Engineers

[48] Varouchakis, E. A., Hristopulos, D. T., Comparison of stochastic and deterministic methods for mapping groundwater level spatial variability in sparsely monitored basins. Environ Monit Assess DOI 10.1007/s10661-012-2527- y

[49] Varouchakis, E.A., Hristopulos, D.T., Karatzas, G.P., Improving kriging of groundwater level data using nonlinear normalizing transformations — a field application, Hydrological Sciences Journal 2012, 57:7, 1404-1419

[50] Varouchakis, E.A., Hristopulos, D.T., Improvement of groundwater level prediction in sparsely gauged basins using physical laws and local geographic features as auxiliary variables, Advances in Water Resources 52 (2013) 34–49

[51] Varouchakis E.A., Geostatistical Analysis and Space-Time Models of Aquifer Levels: Application to Mires Hydrological Basin in the Prefecture of Crete, Dissertation, Technical University of Crete, 2012.

[52] Volpi G, Gambolati G. On the use of a main trend for the kriging technique in hydrology. Adv Water Resour 1978;1:345–9.

[53] Webster R., Oliver M.A., Geostatistics for Enviromental Scientists, John Wiley and Sons Inc, second edition,2009

[54] Xinjie Yu, Mitsuo Gen, Introduction to Evolutionary Algorithms, Springer 2010, DOI 10.1007/978-1- 84996-129-5

[55] Yang F-g, Cao S-y, Xing-nian L, Yang K-j. Design of groundwater level monitoring network with ordinary kriging. J Hydrodyn Ser B 2008;20: 339–46.