

## A decision support tool for optimising groundwater-level monitoring networks using an adaptive genetic algorithm

Antonios Parasyris<sup>a</sup>, Katerina Spanoudaki<sup>a,\*</sup>, Emmanouil A. Varouchakis<sup>b</sup> and Nikolaos A. Kampanis<sup>a</sup>

<sup>a</sup> Foundation for Research and Technology-Hellas, Institute of Applied and Computational Mathematics, Heraklion, Crete, Greece

<sup>b</sup> School of Environmental Engineering, Technical University of Crete, Chania, Greece

\*Corresponding author. E-mail: kspanoudaki@gmail.com

### ABSTRACT

Mapping of the spatial variability of sparse groundwater-level measurements is usually achieved by means of geostatistical methods. This work tackles the problem of deficient sampling of an aquifer, by employing an innovative integer adaptive genetic algorithm (iaGA) coupled with geostatistical modelling by means of ordinary kriging, to optimise the monitoring network. Fitness functions based on three different errors are used for removing a constant number of boreholes from the monitoring network. The developed methodology has been applied to the Mires basin in Crete, Greece. The methodological improvement proposed concerns the adaptive method for the GA, which affects the crossover–mutation fractions depending on the stall parameter, aiming at higher accuracy and faster convergence of the GA. The initial dataset consists of 70 monitoring boreholes and the applied methodology shows that as many as 40 boreholes can be removed, while still retaining an accurate mapping of groundwater levels. The proposed scenario for optimising the monitoring network consists of removing 30 boreholes, in which case the estimated uncertainty is considerably smaller. A sensitivity analysis is conducted to compare the performance of the standard GA with the proposed iaGA. The integrated methodology presented is easily replicable for other areas for efficient monitoring networks design.

**Key words:** adaptive genetic algorithm, geostatistical modelling, groundwater monitoring network optimisation, kriging-based genetic algorithm optimisation

### HIGHLIGHTS

- Development of an innovative adaptive genetic algorithm for optimising groundwater-level monitoring networks.
- Coupling of evolutionary algorithms with geostatistics for monitoring network optimisation.
- Development of a monitoring network design optimisation tool, easily applicable to any area, which considerably reduces sampling efforts, while achieving accurate mapping of groundwater levels.

### INTRODUCTION

Protecting water resources is part of the sixth goal of the United Nations proposed to transform our world in order to promote prosperity while protecting the planet (United Nations 2018). The growing world population as well as a 40% estimated shortfall in freshwater resources by 2030 are leading towards a global water crisis. Groundwater monitoring networks provide essential information for water resources management, especially in areas with significant groundwater exploitation for agricultural and domestic use. Data from such networks are typically used by competent authorities and scientists to validate groundwater flow and contaminant transport models, to assess the response of groundwater levels to pumping, artificial recharge and changing climatic variables and to regulate groundwater exploitation to ensure the sustainability of aquifer resources. Given the high maintenance costs of monitoring networks, the development of tools, which can be used by regulators for efficient network design, is essential. The design of a groundwater monitoring network depends on the spatial and temporal distribution of water levels in the aquifer and the location of potential contaminant sources. These distributions depend on hydrogeological parameters, aquifer characteristics (e.g. confined, unconfined aquifers), physical aquifer boundaries, hydraulic connection between groundwater and surface water bodies, etc. A typical objective for long-term monitoring of groundwater quality is the development of a cost-effective design that adequately characterises a contaminant plume. For long-term monitoring of water levels, the typical objective is the development of a cost-effective network that retains the monitoring boreholes, which contribute to the accurate representation of the spatial variability of the aquifer's groundwater level and excludes boreholes that add little or no beneficial

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (<http://creativecommons.org/licenses/by/4.0/>).

information to groundwater-level mapping of the area. At the same time, network design tools can indicate the expansion of a monitoring network to areas of increased uncertainty for groundwater level or quality variables.

In relevant publications found in the literature, two common heuristic search approaches have been followed for optimising groundwater monitoring networks by identifying the optimum set of observation wells and specifying redundant sampling locations: decision support tools (e.g. Cameron & Hunter 2000; Aziz *et al.* 2002) and mathematical optimisation (e.g. Reed *et al.* 2000; Gangopadhyay *et al.* 2001; Asefa *et al.* 2004; Nunes *et al.* 2004a, 2004b; Kollat & Reed 2005; Wu *et al.* 2005; Theodossiou & Latinopoulos 2006; Li & Hilton 2007; Khan *et al.* 2008; Dhar & Datta 2010; Fisher 2013; Thakur 2015). These approaches have been combined with numerical groundwater flow and contaminant transport simulation models, deterministic (e.g. inverse distance weighting, IDW) or stochastic (e.g. ordinary kriging, OK) interpolation methods and/or statistical analysis to estimate groundwater levels or contaminant concentrations (Li & Hilton 2007). Decision support tools have been applied for improving existing monitoring networks by analysing current and historical groundwater monitoring data. For example, Aziz *et al.* (2002) developed the monitoring and remediation optimisation system (MAROS). MAROS is a decision support tool that uses Delaunay triangulation combined with a ranking rule-based approach, based on spatial data analysis, to reduce the number of sampling locations, and statistical analysis of time series data (trend estimation) to reduce sampling frequency. Cameron & Hunter (2000) developed the geostatistical temporal/spatial optimisation algorithm, which is a site-specific statistical method based on kriging for reducing large monitoring networks. The main disadvantage of decision support tools is that they use manual iterative processes rather than automatic optimisation, and therefore global optimal search and sensitivity analysis under different constraints are not included. Therefore, mathematical optimisation techniques have been adopted much more widely for network design problems.

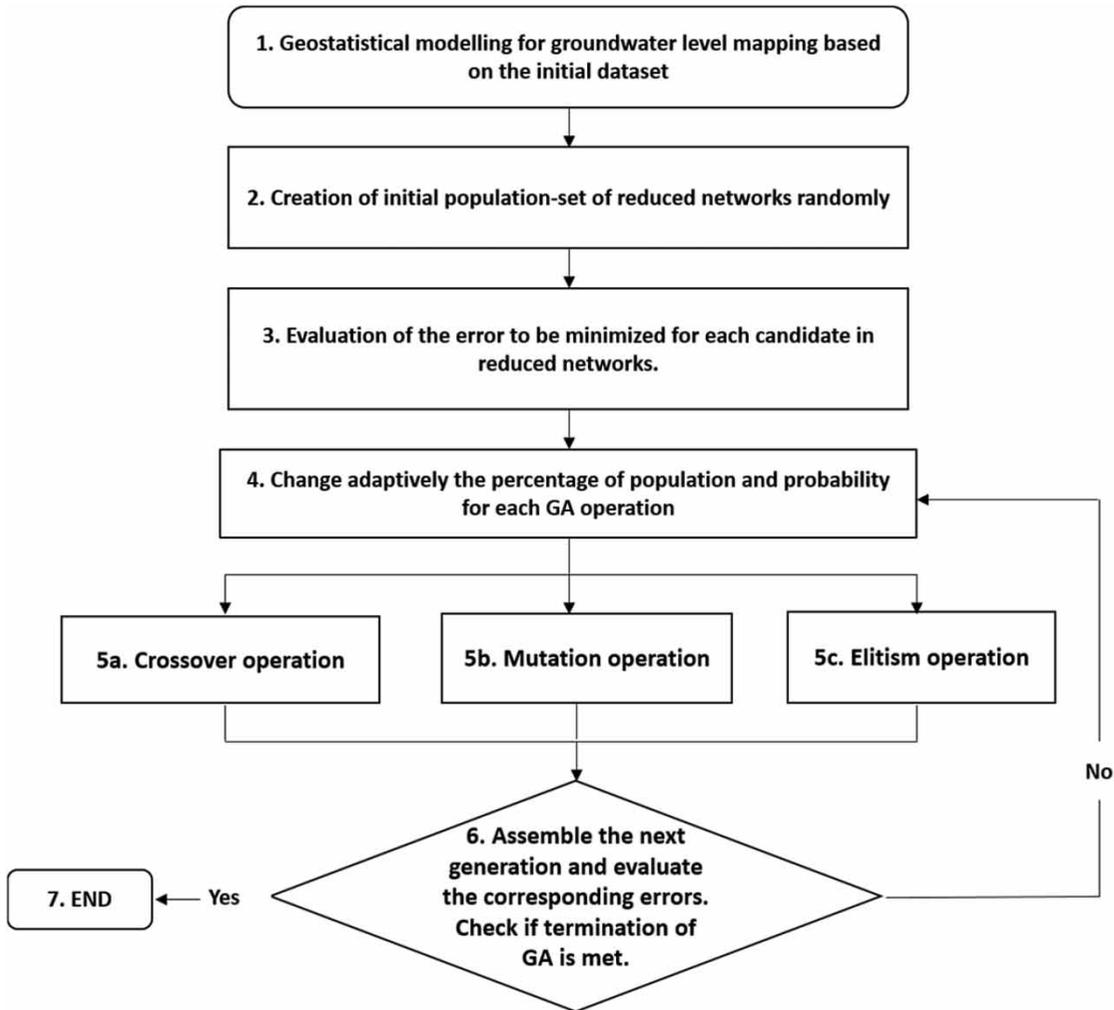
In general, the optimisation of groundwater monitoring networks is a nonlinear combinatorial problem and, therefore, is well suited for heuristic optimisation algorithms, such as genetic algorithms (GAs) (Jiabao *et al.* 2008), simulated annealing (SA), artificial neural networks (ANNs), support vector machines (SVMs) and ant colony optimisation (ACO). For example, Cieniawski *et al.* (1995) optimised groundwater monitoring networks using GAs combined with Monte Carlo simulation. Reed *et al.* (2000) applied IDW and OK combined with GAs. Villas-Boas *et al.* (2017) used ANNs to assess the redundancy of a water quality monitoring network and rank parameters and monitoring locations based on their relevance. Nunes *et al.* (2004a) employed SA to accurately map the spatial variability of groundwater-level networks, while the methodology used in Asefa *et al.* (2004) is based on SVMs and Li & Hilton (2007) applied an ACO algorithm combined with IDW. In subsequent works, GAs have been used in optimal control and monitoring of water and sewage applications (e.g. Johns *et al.* 2014; Soroush & Abedini 2019; Mounce *et al.* 2020) using standard GA methodology and geostatistical algorithms.

In this work, a monitoring network optimisation tool is presented, which couples geostatistical modelling with a GA method. The purpose of the optimisation tool is to determine which boreholes to exclude from the monitoring network if they add little or no beneficial information to groundwater-level mapping of a specific area. Unlike previous relevant investigations, the network optimisation tool presented here uses OK with the recently established non-differentiable Spartan semivariogram for groundwater-level mapping (Varouchakis *et al.* 2012; Varouchakis & Hristopoulos 2013). The geostatistical model has been coupled with an integer adaptive Genetic Algorithm (iaGA) coded in MATLAB (2019a). An improvement made to the classic GA is the change of the mutation and crossover fraction in an adaptive setting with respect to the change of the mean fitness value. This results to a randomness in reproduction, if the solution converges, to avoid local minima or in a more educated reproduction (higher crossover ratio) when there is higher change in the mean fitness value.

The paper is structured as follows: the methodological formulation is described in the 'Methodology' section, which includes the geostatistical modelling formulation and the GA introduction. Along with the introduction of the evolutionary algorithm, the methodology employed to improve the GA and to overcome difficulties in implementation is presented, introducing alterations to MATLAB's optimisation toolbox. In the 'Results and discussion' section, the results obtained, the semivariogram fittings and the residual error estimations are discussed, and the different removal scenarios are analysed, including a sensitivity analysis of the proposed aiGA compared to the classical GA technique. Finally, the 'Conclusions' section summarises the findings and novelties of this work.

## METHODOLOGY

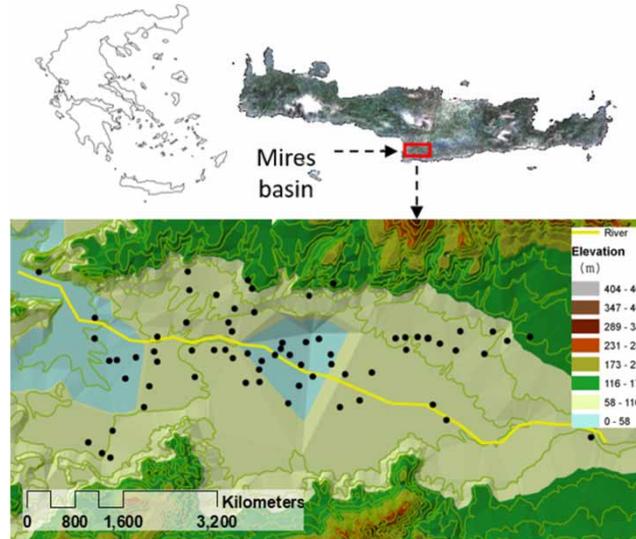
An overview of the methodological approach, coupling geostatistical modelling with GA optimisation, is shown in Figure 1. Geostatistical modelling is employed to estimate the spatial distribution of groundwater levels, or groundwater-level mapping



**Figure 1** | Flow diagram of the methodological approach.

for the area of interest, based on an initial dataset of existing boreholes. Then, a random population of a predetermined number of removals (herein called scenario) is assembled, and the error to be minimised between the reduced population mappings and the target initial population mapping is calculated. Following the standard GA procedure, the population is sorted and split into three groups so that they either undergo crossover, mutation or they are kept unaltered (elitism). An innovative part of the methodological tool is shown at step 4, in which the adaptivity step based on the stall parameter is introduced and affects each GA iteration based on the performance of the current population, as is discussed more in depth in the following section. After the main operations of the GA, the new population is assembled and sorted in order to either terminate with a criterion or repeat the procedure from step 4 using the newest population.

The developed methodology has been applied to the Mires basin in Crete, Greece, an area of high socio-economic and agricultural interest, which suffers from groundwater overexploitation leading to a significant decrease in groundwater levels. The study area is shown in [Figure 2](#) indicating the dense grid of 70 boreholes, which operate in the area for groundwater abstraction and water-level monitoring. Regarding geostatistical modelling, OK with the non-differentiable Spartan semivariogram leads to optimal groundwater-level mapping based on a previous geostatistical study in the area ([Varouchakis et al. 2012](#); [Varouchakis & Hristopoulos 2013](#)). Overall, the Spartan semivariogram resulted in the most accurate groundwater-level estimates followed closely by the power-law model. The optimisation algorithm has been applied to find the set of boreholes whose removal leads to the minimum error between the original water-level mapping using all the available boreholes in the network and the groundwater-level mapping using the reduced borehole network (error is defined as the two-norm of the difference between the original mapping matrix with 70 wells and the mapping matrix of the reduced well network).



**Figure 2** | Mires basin in Crete, Greece, with 70 boreholes where groundwater-level measurements are available.

The solution to the optimisation problem depends on the total number of boreholes that are chosen to be removed, which is an *a priori* problem-specific management decision. To achieve the optimum solution in the minimum possible computational time, a ‘stall generations’ termination criterion was selected.

The choice of an integer GA in MATLAB poses the restriction of adding custom selection and crossover–mutation functions. Therefore, custom population and crossover–mutation–selection functions have been created to set the initial population type to custom and have the ability to change the mutation crossover probability in respect to the convergence of the GA, thus achieving higher accuracy. The use of a GA was necessitated by the sophistication level of the geostatistical tool and the numerous combinations of borehole removal scenarios. It is important to note that a hard search of all the possible scenarios of boreholes/measurements removal from the initial dataset of 70 boreholes would take approximately  $10^8$  years. Additionally, in order to estimate the minimum error between the original mapping and the reduced mapping, the problem would not be easily solved as a typical gradient-based optimisation, as there is no closed form of the function for differentiation. More details on the geostatistical modelling and the GA optimisation are given in the following subsections.

### Geostatistical modelling

It is assumed that the hydraulic head is represented by a spatial random field (SRF), which herein will be denoted by  $Z(\mathbf{s})$ . A prerequisite is a sampled field at measurement points which is denoted by  $Z(\mathbf{s}_i)$  with  $i$  being an index of these points. The goal is the derivation of estimates  $z(\mathbf{s}_0)$  for every point  $\mathbf{s}_0$  in a given rectangular grid. Prior to the application of OK, a normalising transformation has been applied to the available dataset, since OK is an optimal estimator if data follow a multivariate normal distribution and the true semivariogram is known (Clark & Harper 2000). Motivated by the Box-Cox transformation applied to hydrological data (Thyer *et al.* 2002), the modified Box-Cox transformation has been used to distribute the available data into approximately Gaussian, and then, after the OK, a suitable back transformation has been implemented. Details on the OK method and the modified Box-Cox transformation can be found in Varouchakis *et al.* (2012) and Varouchakis & Hristopulos (2013).

Under the second-order stationarity hypothesis, the semivariogram and the covariance function are equivalent. For reasons of convenience, the semivariogram structure is preferred. After the definition of a lag, which is an interval size for the discretisation length of the distances between the measurements, the discrete empirical semivariogram has been defined by using the standard Matheron method-of-moments estimator, which reads to,

$$\hat{g}(\mathbf{r}) = \frac{1}{2N(\mathbf{r})} \sum_{i=1}^{N(\mathbf{r})} \{[z(\mathbf{s}_i) - z(\mathbf{s}_i + \mathbf{r})]^2\} \quad (1)$$

where  $N(\mathbf{r})$  is the number of pairs at lag  $\mathbf{r}$  (Kitanidis 1997).

There were no distinct anisotropies in the dataset as shown in Varouchakis & Hristopulos (2013), so the empirical semi-variogram could be fitted with omnidirectional model functions. The classical semivariogram functions (Table 1) as well as the Spartan family variograms (Varouchakis *et al.* 2012) have been tested. The definition of a theoretical semivariogram is achieved by defining the covariance function between the measurement points  $\gamma_z(\mathbf{r}) = \sigma_z^2 - C_z(\mathbf{h})$ . So, the Spartan covariance function is defined as follows:

$$C_z(\mathbf{h}) = \begin{cases} \frac{\eta_0 e^{-h\beta_2}}{2\pi\sqrt{|\eta_1^2 - 4|}} \left[ \frac{\sin(h\beta_1)}{h\beta_1} \right], & \text{for } |\eta_1| < 2 \\ \frac{\eta_0 e^{-h}}{8\pi}, & \text{for } \eta_1 = 2 \\ \frac{\eta_0(e^{-h\omega_1} - e^{-h\omega_2})}{4\pi(\omega_2 - \omega_1)h\sqrt{|\eta_1^2 - 4|}}, & \text{for } \eta_1 > 2 \end{cases} \quad (2)$$

where  $\eta_0$  is the scale factor,  $\eta_1$  is the rigidity coefficient,  $\beta_{1,2} = |2 \mp \eta_1|^{1/2}/2$ ,  $\omega_{1,2} = (|\eta_1 \mp \Delta|/2)^{1/2}$ ,  $\Delta = |\eta_1^2 - 4|^{1/2}$ ,  $\xi$  is a characteristic length,  $\mathbf{h} = \mathbf{r}/\xi$  is the normalised lag vector and  $h = |\mathbf{h}|$  is the Euclidean norm (Hristopulos 2003, 2020).

### GA optimisation

GAs belong to the family of evolutionary algorithms and use heuristic mechanisms in order to minimise a fitness function (e.g. Holland 1984; Sivananda & Deepa 2008). To justify the need of a heuristic algorithm, other optimisation techniques must be ruled out, since heuristic algorithms are very computationally intense, and do not guarantee the calculation of the global minimum (Yu & Gen 2010). In addition, if the function has multiple minima, the GA is only guaranteed to find or approximate one of them, with no information on the plurality of local minima.

To justify the use of the GA, the irregularity of the function to be minimised is presented in Figure 3, where the output is obtained with the removal of just one borehole. In this simplistic scenario, a trial-and-error technique can be used to find the minimum, which is given when borehole number 68 is removed from the sample. The corresponding value of the output (root-mean-square deviation (RMSD) = 0.3292, explained in the 'Results and discussion' section) is the error one calculates by removing this borehole from the initial mapping. This error function is obviously not differentiable, and no smoothing or closed form can easily be applied in order to find a minimum with other classical methods.

Additionally, in order to compute all the possible reduced network mappings using the current state-of-the-art computers and a trial-and-error technique, one would need roughly  $10^8$  years, e.g. a 30-removal scenario, which renders the problem intractable with a brute force method (Parasyris 2016). To proceed with the proposed GA, the error metrics to be optimised must be defined. Three error metrics have been introduced for the GA minimisation. Their properties and the corresponding results have been compared in order to determine the optimum reduced network for an accurate geostatistical mapping. Two of these errors have been applied by only making an estimate on the missing boreholes; therefore, computations are very fast,

**Table 1** | Theoretical model functions

---

Exponential:  $\gamma_z(\mathbf{r}) = \sigma_z^2 \left[ 1 - \exp\left(-\frac{|\mathbf{r}|}{\xi}\right) \right]$

Gaussian:  $\gamma_z(\mathbf{r}) = \sigma_z^2 \left[ 1 - \exp\left(-\frac{\mathbf{r}^2}{\xi^2}\right) \right]$

Spherical:  $\gamma_z(\mathbf{r}) = \sigma_z^2 \left[ \frac{1.5|\mathbf{r}|}{\xi} - 0.5 \left( \frac{|\mathbf{r}|}{\xi} \right)^3 \vartheta(\xi - |\mathbf{r}|) \right]$ , where  $\{\xi = 0, \text{ if } \xi - |\mathbf{r}| < 0, \xi = 1, \text{ if } \xi - |\mathbf{r}| > 0\}$

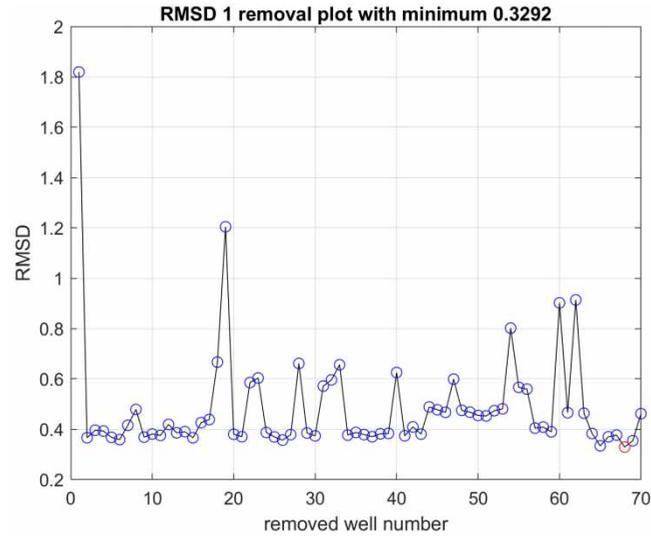
Power – law:  $\gamma_z(\mathbf{r}) = c|\mathbf{r}|^{2H}$ ,  $0 < H < 1$

Linear:  $\gamma_z(\mathbf{r}) = c|\mathbf{r}|$

Matérn:  $\sigma_z^2 \left\{ 1 - \frac{2^{1-\nu}}{\Gamma(\nu)} \left( \frac{|\mathbf{r}|}{\xi} \right)^\nu K_\nu \left( \frac{|\mathbf{r}|}{\xi} \right) \right\}$

---

$\sigma_z^2$  is the variance,  $|\mathbf{r}|$  is the Euclidean norm of the lag vector  $\mathbf{r}$ ,  $\xi$  is the correlation length,  $c$  is the coefficient,  $H$  is the Hurst exponent,  $\nu$  is the smoothness parameter,  $\Gamma(\cdot)$  is the Gamma function and  $K_\nu(\cdot)$  is the modified Bessel function of the second kind of order  $\nu$ .



**Figure 3** | Error function to be minimised with the removal of one borehole.

but generally lack accuracy. In contrast, the third error metric uses the whole prediction matrix; therefore, it is expected to be more accurate but subsume a significant amount of computational time for the heuristic optimisation step. More specifically, the first metric, RMSE (root-mean-square error), is inspired by the leave-one-out cross-validation process. Instead of leaving one borehole out of the computations, a scenario of well removals is designed, and by initially randomly selecting which boreholes to leave out, a reduced dataset is produced. After the geostatistical analysis for the reduced dataset is conducted, the predicted and observed groundwater-level values that are left out of the dataset are compared. To formulate it, we define:

$$e_{\text{RMS}} = \sqrt{\frac{1}{N} \sum_{i=1}^N [z^*(\mathbf{s}_i) - z(\mathbf{s}_i)]^2} \quad (3)$$

where  $N$  is the number of boreholes that have been left out,  $z^*(\mathbf{s}_i)$  is the kriging estimated groundwater level using the remaining  $(n-N)$  dataset and  $z(\mathbf{s}_i)$  is the measured groundwater level at the corresponding point/boreholes that have been removed from the dataset.

The second metric is the Akaike criterion (Akaike 2011), and it has also been calculated at the cross-validation stage; therefore, it is also expected to be optimised by the GA with very little computational cost. The metric that will be referred hereby as Akaike error is defined as follows:

$$I_{\text{AIC}} = n \log \left( \frac{\sum_{i=1}^n (\hat{\gamma}(r_i) - \gamma(r_i))^2}{n} \right) + 2\mu \quad (4)$$

where  $n$  is the lag number,  $\hat{\gamma}(r_i)$  is the value of the experimental semivariogram at lag  $i$ ,  $\gamma(r_i)$  is the value of the theoretical semivariogram at lag  $i$  and  $\mu$  is the number of degrees of freedom that each theoretical semivariogram has. This criterion has been mainly used to compare the semivariograms' fitting efficiency (the lower the better), and it also takes into account the number of semivariogram parameters. Minimising with this metric not only gives an excellent fit to the experimental semivariogram but also gives an indication of which is the best theoretical semivariogram that fits optimally with each reduced network.

The third metric gives the most accurate reduced mapping in comparison to the original mapping, and it is defined to be the RMSD between every point of the discretised SRF. It is defined as follows:

$$e_{\text{RMSD}} = \sqrt{\frac{1}{M} \sum_{i=1}^M [z_{\text{init}}(\mathbf{s}_i) - \hat{z}(\mathbf{s}_i)]^2} \quad (5)$$

where  $M$  is the number of points inside the convex hull in the grid (for the Mires basin application, a  $100 \times 100$  grid was selected),  $z_{\text{init}}(\mathbf{s}_i)$  are the predicted values at the  $M$  points using the original monitoring network and  $\hat{z}(\mathbf{s}_i)$  are the predicted values at the same  $M$  points using the reduced monitoring network.

The basic processes of a GA include (e.g. Yu & Gen 2010):

- the *Creation* of the original population,
- the *Evaluation* of these initial candidates (called parents) using the function to be minimised (called fitness function),
- the *Crossover* operation, which acts on two parents by exchanging genes with each other,
- the *Mutation* operation in which parents that are less well fitted have their genes randomly altered and
- the *Elitism* operation in which parents are transferred unaltered into the next generation.

In this work, the fitness function takes as input a set of removals (boreholes) and has as output one of the errors mentioned above. A sorting occurs, based on the fitness function output. The pivotal step takes place at the stage, where the next population (offspring) is created by using the parents' genes. Every parent is a set of integer indices that correspond to monitoring borehole numbers. This is achieved in three ways: First, the higher percentage of the population undergoes the Crossover operation. To choose which genes will be exchanged, the uniform crossover technique has been selected with a certain crossover probability, initially set to 50% but changed adaptively afterwards. Secondly, a smaller percentage of the population undergoes the Mutation operation. This is beneficial in case the optimal candidate is not included in the initial population, so mutation offspring can search for genes (in this work borehole numbers) that could not be replicated by using the mixing mechanism of Crossover. Lastly, to ensure that the best-fitted candidate is kept in the next generation, the Elitism operation has been employed for a very small percentage of the best-fitted parents. Following the creation of the new population, the GA checks if certain termination criteria have been met; otherwise, it returns to the initial step of Evaluation. The most common termination criterion used is based on the stall generations, which are the generations that have been passed without the best candidate to change. Other termination criteria could be based on the tolerance of the fitness function, a time limit, a total generation limit, etc.

The idea of an adaptive GA is not new and can be traced back to Srinivas & Patnaik (1994), but the methodology by which each author implements adaptivity can differ and be novel. In Pellerin *et al.* (2004), the authors have reviewed the improvements in GA methods, leading to the introduction of an adaptive GA. More precisely, the proposed adaptive GA method alters the crossover and mutation probabilities based on the genetic diversity measure (gdm), which is defined as the ratio between the mean and maximum values of the fitness value for each generation. This way, better convergence times of the GA as well as better fitness values are achieved. The idea is based on the fact that the rate at which mutation occurs needs to grow when the gdm is close to unity, whereas the crossover operation is used more when the gdm is lower than one. The resulting algorithm adjusts the probabilities for each operation accordingly.

Based on the work of Pellerin *et al.* (2004), an alternative version of the adaptive GA developed and presented in this work, which changes these rates and probabilities, is based on the value of the stall parameter. This approach does not necessitate the calculation of any additional metrics and is also simpler to implement than the method of Pellerin *et al.* (2004). Moreover, the adaptive GA proposed in this work shows improved accuracy and convergence speed compared to the standard GA, as shown by the sensitivity analysis conducted, discussed in the 'Results and discussion' section. More specifically, when the stall generations are higher than a given threshold – the chosen value was every 5 stall generations, but this is a problem-based decision – a gradual increase in the mutation probability and an analogous reduction of the crossover probability are introduced. Another innovative feature is that the fraction of the population to undergo crossover and mutation is modified in a similar manner, which contributes towards a higher accuracy and lower computational time of the proposed adaptive GA. Initial values of 80% crossover, 15% mutation and 5% elitism fractions have been chosen in this work. When the stall generations parameter reaches 10, the mutation fraction is increased by 10% whereas the crossover fraction is decreased by 10%.

The procedure continues to alter these fractions until either a fraction threshold of 80% is reached or a better candidate is found that reduces the stall parameter to zero. In the latter case, the GA is designed to return to the initial fraction values. This will result in an increased randomness when a higher stall generation is detected, where the GA is unable to improve its best candidate using the existing population. On the other hand, this variable fraction GA method allows for a higher-than-normal crossover fraction at the initial stage, whereas in fixed fraction GAs, lower crossover might be used to account for the randomness that is not necessary at the initial stage of the GA. The proposed adaptive GA is referred to as an iaGA since the input are vectors with integers (number of boreholes to be removed), which was not available in the software used for the implementation (MATLAB 2019a). To resolve this issue, the GA optimisation tool of MATLAB was customised accordingly by altering appropriately the functions for the creation of initial population, crossover and mutation.

## RESULTS AND DISCUSSION

### Borehole removal scenarios

The semivariogram fitting process for the initial dataset showed that the Spartan semivariogram followed by the power-law (Figure 4) provides the best fits validating the results reported in Varouchakis & Hristopulos (2013). Semivariograms do not reach a sill, denoting that long-distance variations exist in the study area. A comprehensive comparison between other semivariogram methods for the dataset of the Mires basin can be found in Varouchakis & Hristopulos (2013). In Table 2, the optimised parameters for the two model semivariograms used herein (Spartan and power-law) are shown. In this work, we have not focused on a detailed investigation of semivariogram calculation for the initial groundwater-level data mapping, as this is scrutinised in the cited work. The authors conclude that the Spartan and power-law semivariograms show the lowest errors. Therefore, in this study, these two semivariograms have been compared in a reduced monitoring network scenario to observe how these statistical models behave during the coupling with the iaGA. The OK interpolation results regarding the initial dataset by means of the Spartan and power-law variograms are presented in Figures 5 and 6.

Considering that the Spartan and power-law variograms provide the most accurate interpolation results, these variograms have been employed as the basic spatial dependence functions to test the mapping efficiency during the optimisation

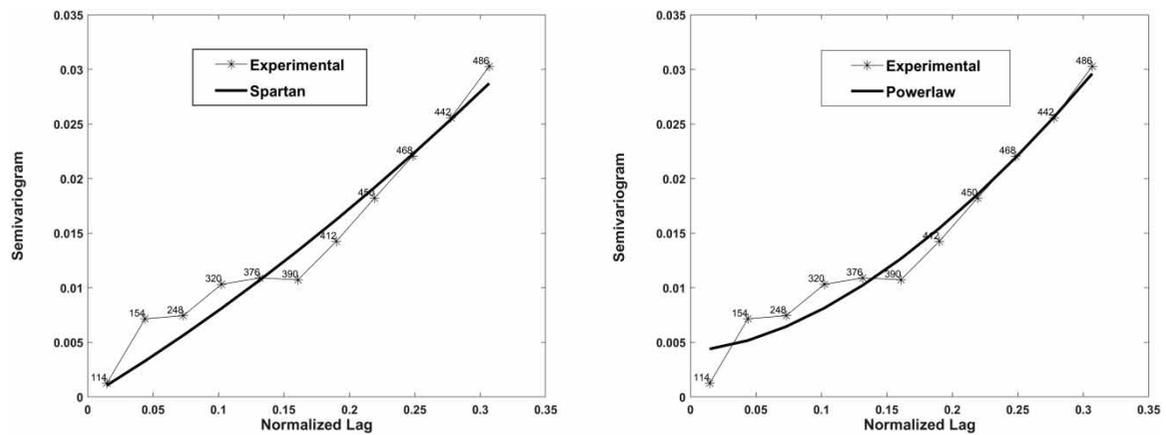
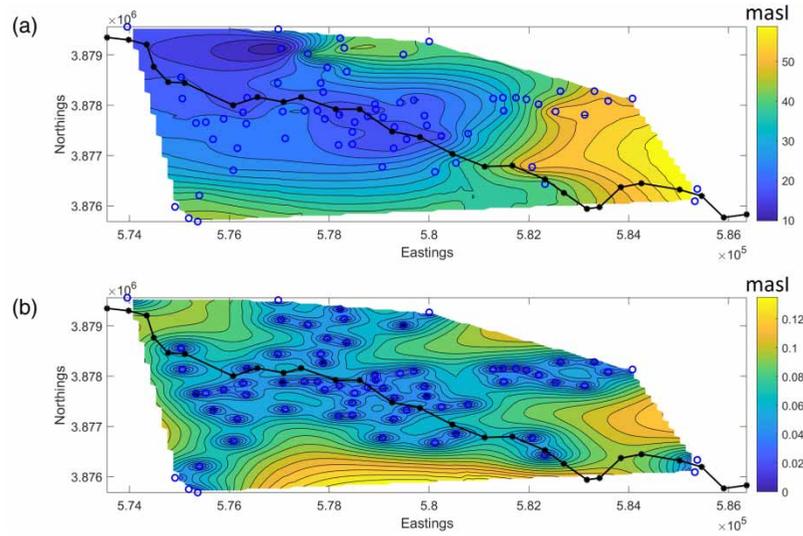


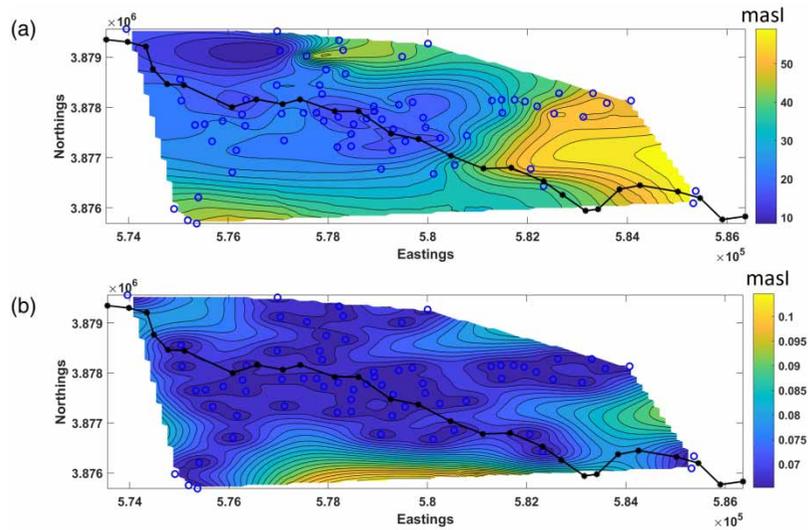
Figure 4 | Fitting of the Spartan and power-law semivariograms on the experimental semivariogram.

Table 2 | Optimised parameters for the two semivariograms used (Spartan and power-law)

Parameters	Methods	
	Spartan	Power-law
$\sigma_z^2$ or $c$	0.2507	0.1903
$\xi$	0.4905	N/A
Nugget	N/A	0.0042
Other parameters	$\eta_1 = -1.9799$	$H = 1.7052$



**Figure 5** | Spartan initial groundwater-level mapping (a) and initial estimated uncertainty (b). Units are in masl (meters above sea level).



**Figure 6** | Power-law initial groundwater-level mapping (a) and initial estimated uncertainty (b). Units are in masl (meters above sea level).

procedure of the monitoring network. The network optimisation scenarios investigated include the removal of 30, 40 and 50 monitoring boreholes. Tables 3 and 4 indicate that removing 30 or 40 boreholes from the initial 70-borehole monitoring network (i.e. removing 43 and 57% of the existing monitoring points) can be considered as efficient monitoring network design scenarios, since for both cases, a high-quality mapping of groundwater levels is retained. The corresponding maps of groundwater-level spatial distribution and of the associated uncertainty, for the reduced monitoring network, using the Spartan variogram function are presented in Figures 7–12. The maps of 30 and 40 removal scenarios are shown here, as the removal of 50 wells resulted in significant errors and is not considered as an effective design solution for the physical problem investigated. Results of this study are comparable to the findings of Fisher (2013) for the optimisation of a groundwater-level monitoring network in the Eastern Snake River Plain Aquifer. Fisher (2013) concluded excluding 12% (considered a safe scenario) or 23% (considered a moderate scenario) of existing wells in the monitoring network. Differences in findings could be owed to the sample variability and density, or the geometry and scale of the areas considered.

Tables 3 and 4 depict the minimum of the errors that the iaGA recorded for the different metrics as well as the deviation from the initial mapping when RMSD is not minimised using the iaGA. This essentially quantifies the deviation of the reduced

**Table 3** | Spartan optimisation errors

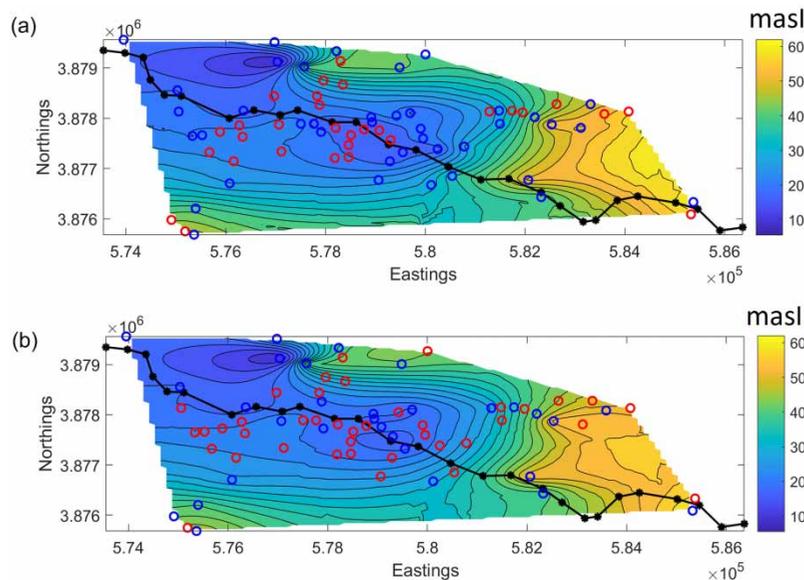
	Scenarios		
	30	40	50
RMSD optimisation	0.5671	0.8250	1.3542
RMSE optimisation	6.8550	14.2811	17.9336
Corresponding RMSDs	0.9528	0.9621	1.1672
Akaike optimisation	-158.3399	-165.3199	-192.8193
Corresponding RMSDs	2.2906	4.9930	8.6190

Smallest possible value for the RMSD, RMSE and Akaike. The corresponding RMSD error is given for the RMSE and Akaike optimisations.

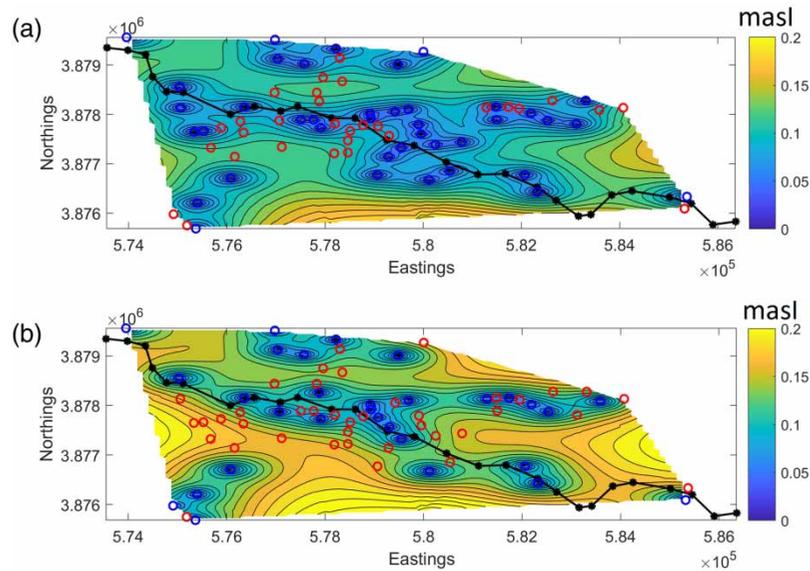
**Table 4** | Power-law optimisation errors

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

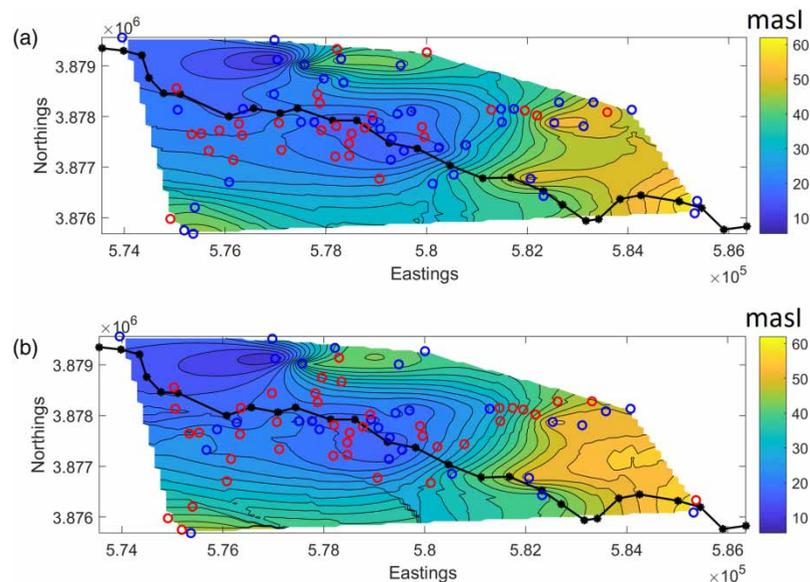
Smallest possible value for the Root Mean Squared Deviation (RMSD), Root Mean Squared Error (RMSE) and Akaike. The corresponding RMSD error is given for the RMSE and Akaike optimisation.

**Figure 7** | Groundwater-level mappings with the Spartan variogram; optimisation based on the RMSD criterion for the removal scenario of 30 (a) and 40 (b) boreholes, respectively. Units are in masl (meters above sea level).

network groundwater-level mapping from the original network mapping using RMSE and Akaike criterion. Most of the errors in Table 3 (Spartan) are lower than the corresponding ones in Table 4 (power-law), which indicates that the Spartan variogram requires less data points for a quality estimation of the spatial distribution of groundwater levels for the specific dataset. Another observation is that the iaGA using the Akaike optimisation can achieve lower error values, as the number of removed



**Figure 8** | Uncertainty estimation mappings with Spartan variogram; optimisation based on the RMSD criterion for the removal scenario of 30 (a) and 40 (b) boreholes, respectively. Units are in masl (meters above sea level).

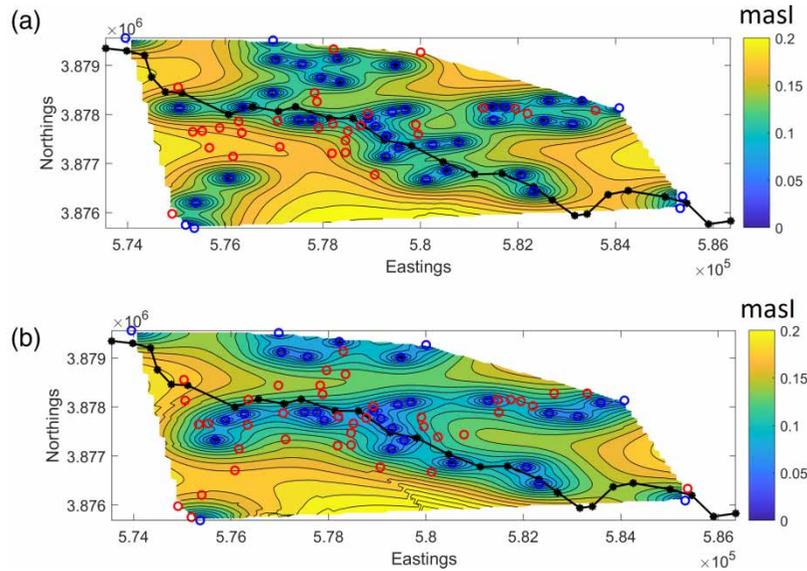


**Figure 9** | Groundwater-level mappings with Spartan variogram; optimisation based on the RMSE criterion for the removal scenario of 30 (a) and 40 (b) boreholes, respectively. Units are in masl (meters above sea level).

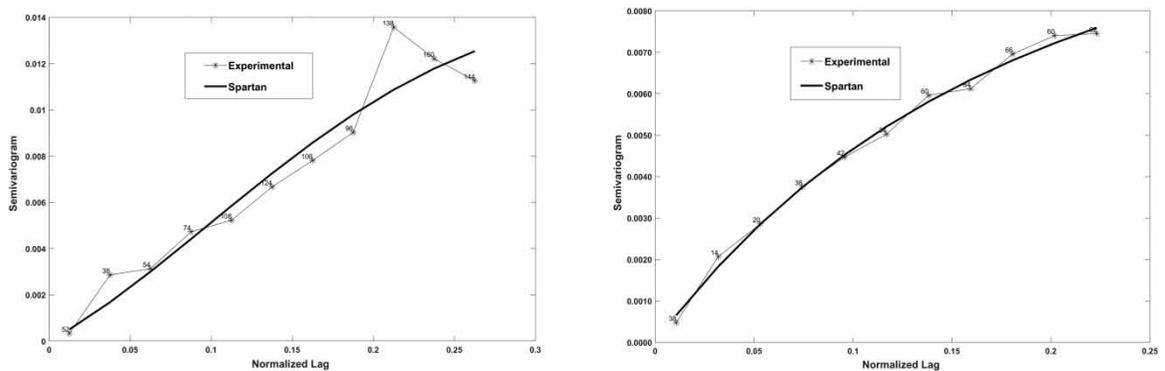
monitoring boreholes is increased. This does not account for a better overall solution as it is observed that RMSD increases significantly, but merely to a better fit of the semivariogram when a reduced number of data points is used.

Another observation is that in the cases with low RMSD, RMSE values (which results in a similar mapping to the original), the kriging variance increases. This is to be expected, since fewer measurements are used for the same prediction domain. Furthermore, when an optimisation with respect to Akaike error is implemented to obtain the monitoring network that provides the best variogram fit (Figures 11 and 12), a smoother distribution formed by the remaining monitoring stations is observed. However, this does not lead to improved interpolation, since the corresponding RMSD and RMSE errors increase.

Regarding the behaviour of the iaGA in relation to the specific attributes of boreholes that were finally removed from the monitoring network, it is depicted in Figures 7 and 9 that using RMSD and RMSE metrics for the optimisation, removed



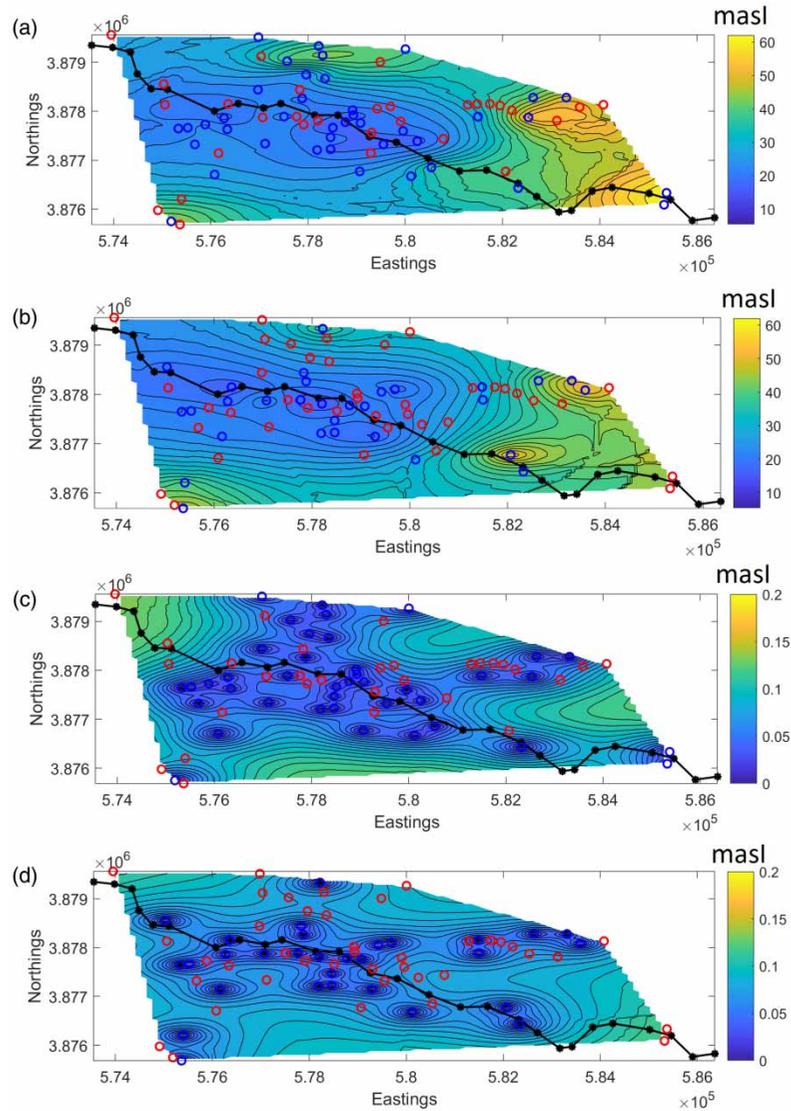
**Figure 10** | Uncertainty estimation mappings with Spartan variogram; optimisation based on the RMSE criterion for the removal scenario of 30 (a) and 40 (b) boreholes, respectively. Units are in masl (meters above sea level).



**Figure 11** | Variogram fittings with the best iaGA solution using the Spartan variogram; removal of 30 (left) and 40 (right) boreholes, respectively, based on the Akaike criterion.

boreholes constitute monitoring points that not only from areas with high sampling density, but also showing little variability between their values. For example, borehole number 9, which is the one historically exhibiting the lowest groundwater-level values, in the northeast of the aquifer, was not removed in the cases of network optimisation with RMSE and RMSD metrics. For the Akaike error minimisation case, extreme groundwater-level values, as those for borehole number 9, are in fact excluded from the optimised monitoring network, and a smoother mapping is created as a result. This should be taken into account when designing a cost-efficient, reliable monitoring network, whether minimising with respect to the best variogram fit is appropriate for the physical problem at hand.

In Tables 5–7, a sensitivity analysis of the iaGA to the population size and the termination criterion stall parameter is presented. The sensitivity analysis has been performed for the case of removing 30 monitoring boreholes, as it has been concluded that this network design scenario is the most cost-efficient, while still retaining an accurate groundwater-level mapping and introducing the least uncertainty. The sensitivity analysis showed that the total time of the RMSD optimisation is considerably high in comparison to the RMSE and Akaike methods. This was expected, as in the latter cases, the fitness function did not include the OK computation on the whole computational grid. The weak dependence of the results on these parameters (population/stall) is also shown, with optimal values found around the 100–200 population size and around 20–40 stall parameters, always considering the computational cost, which increases as these two parameters increase. This



**Figure 12** | Groundwater-level mappings (a,b) and uncertainty obtained (c,d) with the Spartan variogram fittings of Figure 11; Akaike criterion minimisation of 30 (a,c) and 40 (b,d) boreholes, respectively. Units are in masl (meters above sea level).

**Table 5** | Spartan optimisation with RMSD sensitivity analysis on the population number and on the stall termination criterion. Values in bold indicate a comparably low RMSD in combination with a low computational cost of the iGA

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

**Table 6** | Spartan optimisation with Root Mean Squared Error (RMSE) sensitivity analysis on the population number and on the stall termination criterion. Values in bold indicate the lowest RMSE values in combination with a low computational cost of the iGA

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

**Table 7** | Spartan optimisation with Akaike sensitivity analysis on the population number and on the stall termination criterion. Values in bold indicate the lowest Akaike values in combination with a low computational cost of the iGA

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

**Table 8** | Sensitivity analysis comparing the standard GA with the adaptive and semi-adaptive GA

	Generations	Fvals	RMSE
Adaptive GA	118	5,950	8.042
Semi-adaptive GA	103	5,200	9.004
Standard GA	94	4,750	10.27

is not a strict rule, as there is some randomness in the iaGA process, but is an observable trend which is to be expected, and as such, it verifies the hypothesis for the advantages of the iaGA.

To recommend a Population and Stall parameter from these tests, one has to take into account the total time that the algorithm needed in accordance with the improvement of the optimised value. In Table 5, it is clearly shown that the RMSD did not decrease significantly after the value 0.5785 was reached and so one can prefer the 100 population–40 stall choice, since the computational time increased considerably for higher population and stall combinations. With regard to the optimisation of the RMSE depicted in Table 6, one can choose the 200 population–20 stall combination that required just 509 s and returned an error value of 6.855. Lastly, from Table 7 it can be concluded that the 100 population–60 stall choice of parameters was the best one amongst the ones tried, since in just 595 s, the lowest value amongst the analysis was reached.

The increase in computational time for the RMSD optimisation (Table 5) should also be noted, in contrast to the RMSE and Akaike optimisations presented in Tables 6 and 7. Hence, if computational time is of essence, then the RMSE error, which also shows a low corresponding RMSD compared to the Akaike optimisation (see Tables 3–4), should be preferred.

### iaGA sensitivity analysis

A sensitivity analysis has been conducted to compare the iaGA proposed in this work, with the standard GA method, for the 30-borehole removal scenario, using the RMSE error. In the standard GA, a 0.85 crossover fraction is being used uniformly and a 0.1 fraction was chosen as the mutation population. The remaining 0.05 fraction of the population is used for elitism. In the iaGA case, it is proposed to start with a 0.85 crossover fraction, and in every 5 stalled generations, the percentage is reduced by 10%. A semi-adaptive case has also been implemented for comparison purposes, in which crossover fraction begins with 0.85 and it drops to 0.5 in the case that the best candidate is being kept the same over 10 iterations/generations (stall = 10). The results presented in Table 8 were obtained running a GA with a population of 50 candidates and a stall termination criterion of 20 generations for benchmarking purposes. The results presented in Table 8 indicate that as the adaptivity level increases from standard GA to semi-adaptive and then to the fully adaptive case, the RMSE error drops from 10.27 to 8.042. On the other hand, more generations and function evaluations ( $F_{vals}$ ) are needed, which agrees with the iaGA rationale, in which the crossover fraction is decreased as the stall generations increase, and the possibility of finding another optimum is increased. Hence, the stall generations will drop again to zero once the new minimum has been found and more  $F_{vals}$  are needed overall.

## CONCLUSIONS

In this work, the groundwater monitoring network of the Mires basin was optimised based on the Spartan variogram function in terms of OK interpolation and an adaptive optimisation algorithm in terms of RMSE/RMSD error. Results show that around half of the monitoring boreholes can be removed from the original dataset with a relatively small error. More specifically, the application of the network optimisation tool to Mires indicates that as many as 40 monitoring boreholes out of 70 can be removed with a small increase in the proposed error metrics. However, this study suggests that the optimal, cost-efficient and reliable monitoring network design scenario is to exclude 30 boreholes from the network, which results both in high-quality groundwater-level mapping, compared to the mapping produced for the initial 70-borehole network, and also introduces low uncertainty in the estimations. The results indicate the robustness of the network optimisation tool: boreholes were removed from high-density monitoring areas while preserving the spatial pattern of the original groundwater-level map. New adaptivity methods were considered here for the GA, and the newly established Spartan variogram was used, which is proved to produce superior results for the given basin. The error metric that provided the most accurate results in terms of the optimisation procedure was RMSD, which also provided improved uncertainty estimations. On the other hand, RMSE gave comparable results to the RMSD, with the added benefit of a lower computational cost. Lastly, the iaGA sensitivity analysis showed improved results compared to the standard GA approach, providing thus a useful optimisation tool.

## CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

## DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

## REFERENCES

- Akaike, H. 2011 *Akaike's information criterion*. In: *International Encyclopedia of Statistical Science* (Lovric, M., ed.). Springer, Berlin, Heidelberg. doi:10.1007/978-3-642-04898-2\_110.
- Asefa, T., Kemblowski, M. W., Urroz, G., McKee, M. & Khalil, A. 2004 *Support vectors-based groundwater head observation networks design*. *Water Resources Research* **40**, W11509. doi:10.1029/2004WR003304.
- Aziz, J., Ling, M. & Rifai, H. S. 2002 *MAROS: a decision support system for optimising monitoring plans*. *Ground Water* **41** (3), 355–367. doi:10.1111/j.1745-6584.2003.tb02605.x.
- Cameron, K. & Hunter, P. 2000 *Optimisation of LTM Networks Using GTS: Statistical Approaches to Spatial and Temporal Redundancy*. Air Force Center for Environmental Excellence, Brooks AFB, TX.

- Cieniawski, S. E., Eheart, J. W. & Ranjithan, S. 1995 Using genetic algorithms to solve a multiobjective groundwater monitoring problem. *Water Resources Research* **31** (2), 399e409.
- Clark, I. & Harper, W. V. 2000 *Practical Geostatistics 2000*. Ecosse North America Llc, Columbus, OH.
- Dhar, A. & Datta, B. 2010 Logic-based design of groundwater monitoring network for redundancy reduction. *Journal of Water Resources Planning and Management* **136** (1), 88–94. [http://dx.doi.org/10.1061/\(ASCE\)0733-9496\(2010\)136:1\(88\)](http://dx.doi.org/10.1061/(ASCE)0733-9496(2010)136:1(88)).
- Fisher, J. C. 2013 *Optimisation of Water-Level Monitoring Networks in the Eastern Snake River Plain Aquifer Using a Kriging-Based Genetic Algorithm Method*. U.S. Geological Survey Scientific Investigations Report 2013-5120 (DOE/ID-22224), 74 p. Available from: <http://pubs.usgs.gov/sir/2013/5120/>.
- Gangopadhyay, S., Gupta, A. D. & Nachabe, M. H. 2001 Evaluation of ground water monitoring network by principal component analysis. *Ground Water* **39** (2), 181–191. doi:10.1111/j.1745-6584.2001.tb02299.x.
- Holland, J. H. 1984 Genetic algorithms and adaptation. In: *Adaptive Control of Ill-Defined Systems., NATO Conference Series (II Systems Science)*, vol. 16. (Selfridge, O. G., Rissland, E. L. & Arbib, M. A., eds). Springer, Boston, MA. doi:10.1007/978-1-4684-8941-5\_21.
- Hristopulos, D. T. 2003 Spartan Gibbs random field models for geostatistical applications. *SIAM Journal on Scientific Computing* **24** (6), 2125–2162. <https://doi.org/10.1137/S106482750240265X>.
- Hristopulos, D. T. 2020 *Random Fields for Spatial Data Modeling*. Springer/Nature, Dordrecht, The Netherlands.
- Jiabao, G., Elcin, M. & Aral, M. 2008 Genetic algorithm for constrained optimization models and its application in groundwater resources management. *Journal of Water Resources Planning and Management* **134** (1), 64–72.
- Johns, M., Keedwell, E. & Savic, D. 2014 Adaptive locally constrained genetic algorithm for least-cost water distribution network design. *Journal of Hydroinformatics* **16** (2), 288–301. <https://doi.org/10.2166/hydro.2013.218>.
- Khan, S., Chen, H. F. & Rana, T. 2008 Optimising ground water observation networks in irrigation areas using principal component analysis. *Ground Water Monitoring and Remediation* **28** (3), 93–100. doi:10.1111/j.1745-6592.2008.00204.x.
- Kitanidis, P. K. 1997 *Introduction to Geostatistics: Applications to Hydrogeology*. Cambridge University Press, New York, 249 pp, ISBN: 9780521587471.
- Kollat, J. B. & Reed, P. M. 2005 Comparing state-of-the-art evolutionary multi-objective algorithms for long-term groundwater monitoring design. *Advances in Water Resources* **29** (6), 792–807. <http://dx.doi.org/10.1016/j.advwatres.2005.07.010>.
- Li, Y. & Hilton, A. B. C. 2007 Optimal groundwater monitoring design using an ant colony optimisation paradigm. *Environmental Modeling and Software* **22** (1), 110–116. <http://dx.doi.org/10.1016/j.envsoft.2006.05.023>.
- MATLAB 2019a The MathWorks, Inc., Natick, MA
- Mounce, S. R., Shepherd, W., Ostojin, S., Abdel-Aal, M., Schellart, A. N. A., Shucksmith, J. D. & Tait, S. J. 2020 Optimisation of a fuzzy logic-based local real-time control system for mitigation of sewer flooding using genetic algorithms. *Journal of Hydroinformatics* **22** (2), 281–295. <https://doi.org/10.2166/hydro.2019.058>.
- Nunes, L. M., Cunha, M. C. & Ribeiro, L. 2004a Groundwater monitoring network optimisation with redundancy reduction. *Journal of Water Resources Planning and Management* **130** (1), 33–43. [http://dx.doi.org/10.1061/\(ASCE\)0733-9496\(2004\)130:1\(33\)](http://dx.doi.org/10.1061/(ASCE)0733-9496(2004)130:1(33)).
- Nunes, L. M., Paralta, E., Cunha, M. C. & Ribeiro, L. 2004b Groundwater nitrate monitoring network optimisation with missing data. *Water Resources Research* **40**, W02406. 18 p. doi:10.1029/2003WR002469.
- Parasyris, A. 2016 *Groundwater Monitoring Networks Management Using Genetic Algorithms*. Master Thesis. Available from: <https://elocus.lib.uoc.gr/dlib/3/f/8/metadata-dlib-1482308368-41844-2033.tkl#>.
- Pellerin, E., Pigeon, L. & Delisle, S. 2004 Self-adaptive parameters in genetic algorithms. In: *Data Mining and Knowledge Discovery – DATAMINE*. Vol. 5433, pp. 53–64. doi:10.1117/12.542156.
- Reed, P., Minsker, B. & Valocchi, A. J. 2000 Cost-effective long-term groundwater monitoring design using a genetic algorithm and global mass interpolation. *Water Resources Research* **36** (12), 3751–3741. doi:10.1029/2000WR900232.
- Sivananda, S. N. & Deepa, S. N. 2008 *Introduction to Genetic Algorithms*. Springer-Verlag, Berlin, Heidelberg. ISBN 978-3-540-73189-4.
- Soroush, F. & Abedini, M. 2019 Optimal selection of number and location of pressure sensors in water distribution systems using geostatistical tools coupled with genetic algorithm. *Journal of Hydroinformatics* **21** (6), 1030–1047.
- Srinivas, M. & Patnaik, L. M. 1994 Adaptive probabilities of crossover and mutation in genetic algorithms. *IEEE Transactions on Systems, man and Cybernetics* **24** (4), 656.
- Thakur, J. K. 2015 Optimising groundwater monitoring networks using integrated statistical and geostatistical approaches. *Hydrology* **2** (3), 148–175. doi:10.3390/hydrology2030148.
- Theodossiou, N. & Latinopoulos, P. 2006 Evaluation and optimisation of groundwater observation networks using the Kriging methodology. *Environmental Modelling & Software* **21** (7), 991–1000. <http://dx.doi.org/10.1016/j.envsoft.2005.05.001>.
- Thyer, M., Kuczera, G. & Wang, Q. J. 2002 Quantifying parameter uncertainty in stochastic models using the Box-Cox transformation. *Journal of Hydrology* **265** (1-4), 246–257. [https://doi.org/10.1016/S0022-1694\(02\)00113-0](https://doi.org/10.1016/S0022-1694(02)00113-0).
- United Nations (UN) 2018 *Sustainable Development Goals ODS 6*. Available from: <https://www.un.org/sustainabledevelopment/water-and-sanitation/>.
- Varouchakis, E. A., Hristopulos, D. T. & Karatzas, G. P. 2012 Improving kriging of groundwater level data using nonlinear normalizing transformations-a field application. *Hydrological Sciences Journal* **57** (7), 1404–1419. doi:10.1080/02626667.2012.717174.
- Varouchakis, E. A. & Hristopulos, D. T. 2013 Comparison of stochastic and deterministic methods for mapping groundwater level spatial variability in sparsely monitored basins. *Environmental Monitoring and Assessment* **185**, 1–19. doi:10.1007/s10661-012-2527-y.

- Villas-Boas, M. D., Olivera, F. & de Azevedo, J. P. S. 2017 Assessment of the water quality monitoring network of the Piabanha River experimental watersheds in Rio de Janeiro, Brazil, using autoassociative neural networks. *Environmental Monitoring and Assessment* **189**, 439. <https://doi.org/10.1007/s10661-017-6134-9>.
- Wu, J., Zheng, C. & Chien, C. C. 2005 Cost-effective sampling network design for contaminant plume monitoring under general hydro geological conditions. *Journal of Contaminant Hydrology* **77** (1–2), 41–65. <http://dx.doi.org/10.1016/j.jconhyd.2004.11.006>.
- Yu, X. & Gen, M. 2010 *Introduction to Evolutionary Algorithms*. Springer. doi:10.1007/978-1-84996-129-5.

First received 4 April 2021; accepted in revised form 8 July 2021. Available online 28 July 2021