

Article



# Design of Groundwater Level Monitoring Networks for Maximum Data Acquisition at Minimum Travel Cost

Juana Cázares Escareño, Hugo Enrique Júnez-Ferreira \*, Julián González-Trinidad, Carlos Bautista-Capetillo and Cruz Octavio Robles Rovelo \*

> Doctorado en Ciencias de la Ingeniería, Universidad Autónoma de Zacatecas, Campus UAZ Siglo XXI, Carretera Zacatecas-Guadalajara Km. 6, Ejido La Escondida, Zacatecas C.P. 98160, Mexico; juana.cazares.e@uaz.edu.mx (J.C.E.); jgonza@uaz.edu.mx (J.G.-T.); baucap@uaz.edu.mx (C.B.-C.)

\* Correspondence: hejunez@uaz.edu.mx (H.E.J.-F.); octavio.robles@uaz.edu.mx (C.O.R.R.);

Tel.: +52-492-108-21-70 (H.E.J.-F.)

**Abstract:** Groundwater monitoring networks represent the main source of information about water levels and water quality within aquifers. In this paper, a method is proposed for the optimal design of monitoring networks to obtain groundwater-level data of high spatial relevance at a low cost. It uses the estimate error variance reduction obtained with the static Kalman filter as optimization criteria, while simultaneously evaluating the optimal routes to follow through the traveling salesman problem. It was tested for a network of 49 wells in the Calera aquifer in Zacatecas, Mexico. The study area was divided into three zones, and one working day (8 h) was taken to visit each one, with an average speed of 40 km/h and a sampling time of 0.5 h. An optimal network of 26 wells was obtained with the proposal, while 21 wells should be monitored if the optimal routing is neglected. The average standard error using 49 wells of the original network was 35.01 m, an error of 38.35 m was obtained for 21 wells (without optimal routing) and 38.36 m with the 26 wells selected using the proposal. However, the latter produce estimates closer to those obtained with the 49 wells. Following the proposal, more field data can be acquired, reducing costs.

**Keywords:** groundwater; monitoring networks; geostatistics; Kalman filter; the traveling salesman problem

# 1. Introduction

Groundwater represents 96.3% of the total freshwater on Earth. It is the principal source of water, and necessary for the survival and development of humanity. Around 69% of the extracted groundwater is used for agricultural activities, 19% is required by the industrial sector, and the remaining 12% is used for domestic purposes [1,2]. According to the National Water Commission (CONAGUA in Spanish) of the Mexican government, the water that is available to exploit without altering its natural balance is 451,585 hm<sup>3</sup>; 39% of it corresponds to groundwater, while 61% corresponds to surface waters [3].

For adequate groundwater management, it is necessary to implement monitoring schemes in a set of piezometers or wells that allow for spatial and temporal information on the quality and levels of groundwater to be obtained by carrying out field campaigns. These networks represent the main source of information needed to make inferences about the hydrological behavior of water in the subsoil [4], which is of vital importance since the extraction of considerable volumes of groundwater has caused adverse effects on water levels, as well as quality [5].

The groundwater data collected in space and time are useful for the construction of numerical models that allow for a simulation of the evolution of groundwater levels and quality [6]. These models represent robust tools supporting water management policies. There are monitoring networks in several parts of the world at present, where the level

**Citation:** Cázares Escareño, J.; Júnez-Ferreira, H.E.; González-Trinidad, J.; Bautista-Capetillo, C.F.; Robles Rovelo, C.O. Design of Groundwater Level Monitoring Networks for Maximum Data Acquisition at Minimum Travel Cost. *Water* **2022**, *14*, 1209. https://doi.org/10.3390/w14081209

Academic Editors: Rui Hu, Linwei Hu and Zhi Dou

Received: 4 March 2022 Accepted: 7 April 2022 Published: 9 April 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/). and quality of groundwater and springs are measured at the regional level. The processing and representation of this information is carried out in Geographic Information Systems (GIS).

In many regions, the adequate management and assessment of water is difficult to achieve due to the limited monitoring the groundwater. To overcome this situation, the International Center for the Assessment of Groundwater Resources (IGRAC) belonging to the United Nations Educational, Scientific and Cultural Organization (UNESCO) founded the Global Groundwater Monitoring Network (GGMN). This network is based on a catalog of 166 parameters/variables included in a GIS, with the aim of facilitating access and updating information to gaining broader knowledge about the functioning of water in subsoil with the aim of its sustainable management [7].

In Mexico, a total of 1118 groundwater monitoring stations were reported in 2017 (1096 of them were wells). These were used for the surveillance of the 653 administrative aquifers defined by CONAGUA. The annual monitoring of groundwater quality has been carried out since 2005, especially in the central and northern regions, due to its relevant role in the country's economic activities. The data obtained from this monitoring program were used to evaluate compliance with the Official Mexican Standard (NOM-127-SSA1-1994), which establishes the permissible limits of groundwater quality parameters for human consumption and agricultural purposes [3]. These networks must be designed by following the criteria that allow for the greatest amount of information to be obtained at the lowest possible cost. Accordingly, several researchers have developed methods for the optimal design of monitoring networks by looking at one or several variables, in space or space-time [8,9]. Some designs aim to minimize the error in the estimation of the hydraulic head in space and over time with the lowest possible number of wells, to reduce the cost of exploration. The estimation method that is usually used is kriging, although other works propose the use of the Kalman filter [10,11]. Bierkens et al. [12] use an autoregressive exogenous (ARX) model to estimate changes in groundwater levels. Geostatistical techniques are used to calculate the ARX parameters at unmonitored sites. The regionalized model was incorporated into a spatiotemporal Kalman filter for the optimal prediction of water level variation.

Different methods for the optimal design of groundwater quality and groundwaterlevel monitoring networks used an extended, non-dominated classification genetic algorithm in the optimization [9,10]. Luo et al. [13] sought to optimize a long-term monitoring network of a contaminant concentration where there is a scarcity of hydraulic conductivity data through a method that incorporates total sampling costs and estimation errors for a mass contaminant. Several researchers designed optimal monitoring networks for the study and treatment of groundwater contamination [14–18].

Other works were developed for the estimation and the optimal design of groundwater-level monitoring networks using artificial neural networks [19], or multicriteria analysis implemented in a geographic information system (GIS) [20]. However, the most popular approach for the design of groundwater-level monitoring networks (GLMN) is the application of geostatistical interpolation techniques [4,21–24].

On the other hand, few works have considered the routes that should be followed in sampling campaigns to obtain field data. One alternative is to solve the traveling salesman problem (TSP) to optimize the route and visit several sites of interest [25–29]. Nunes et al. [30] proposed optimal well subsets that will form part of a groundwater-level monitoring network, incorporating the TSP. They built an objective function (OF) that evaluates the reduction in the spatial variance, temporal redundancy and exploration costs (travel and monitoring) for a previously defined fixed number of wells. Simulated annealing (SA) was used to optimize the OF. The problem described is of a combinatorial type and involves a cost–benefit analysis, which guarantees quality information at low cost. With this method, the selected wells vary depending on the size of the network that was set prior to optimization.

The aim of this paper is to present a new method for the design of groundwater-level monitoring networks, to select the maximum number of wells that can be visited with a predefined budget, minimizing travel distances and maximizing the amount of information that can be provided. The objective function considers both the order of priority assigned to each well, according to the reduction in the estimate error variances of the groundwater levels, obtained by means of the static Kalman filter and the order of priority from an analysis of the optimal route, solving the TSP. This proposal was developed based on [30], which incorporates exploration times in the objective function, and [31], which employs a Kalman filter-based method to evaluate the reduction in the total estimate error variance.

# 2. Materials and Methods

# 2.1. Location of the Study Area

The method was tested to redesign the Calera aquifer monitoring network located in Zacatecas, Mexico, between parallels 22°41′ and 23°24′ north latitude and meridians 102°33′ and 103°01′ west longitude, with an area of 2226 km² (Figure 1). The municipalities within the aquifer are: Gral. Enrique Estrada and Morelos, and part of the municipalities of Calera, Fresnillo, Pánuco, Veta Grande and Zacatecas. Most of the population lives in the municipalities of Morelos, Calera, Gral. Enrique Estrada and Fresnillo [32], with a total of 306,142 inhabitants [33].



Figure 1. State of Zacatecas, Mexico, and location of the study area (Calera aquifer).

According to the 2010 census in Mexico, within the Calera aquifer, there are 1417 wells, from which a total volume of 176.5 hm<sup>3</sup> of water is extracted annually. Of this, 159.2 hm<sup>3</sup> is used in the agricultural sector, 11.1 hm<sup>3</sup> for drinking water supply, 1.1 hm<sup>3</sup> for livestock and domestic use, and 5.1 hm<sup>3</sup> for industrial use. The Calera aquifer is of an unconfined, heterogeneous and anisotropic type. Its thickness is approximately 400 m, the

hydraulic conductivity values vary from 0.22 to 2.2 m/day, and it has an average specific yield of 0.13. Geologically speaking, its upper layer largely consists of alluvial material, conglomerate and rhyolite–acid tuff. The underlying layers are composed of volcanic (acid tuffs, rhyolites and ignimbrites) and sedimentary (limestone, shale and sandstone) rocks. In the area of interest, a semidry climate was identified. The rainy season occurs more frequently in the summer. For the 1980–2009 period, the mean annual values for rainfall, temperature, and potential evaporation were 425 mm, 16.3 °C, and 2263 mm, respectively [32]. The most important crops in the region are green chili, beans, corn for grain, onion, garlic and alfalfa, and they are produced with irrigated agriculture, which occupies four times less surface than rainfed agriculture [34].

#### 2.1.1. Existing Groundwater-Level Monitoring Network

There are no piezometers built within the Calera aquifer, so the monitoring network is made up of abandoned wells enabled for this purpose; the existing monitoring network (MN) comprises a total of 49 wells.

The groundwater levels in these wells are measured annually with the use of water level meters. Additionally, water samples are obtained to analyze their quality in the laboratory. Since the monitoring is carried out in this way, a design that considers ways to reduce the costs in the acquisition of piezometric data is needed. For the redesign, groundwater-level data for the year 2017 were used, since this was the most sampled year in recent times. These data were provided by the Department of Groundwater of CONA-GUA in Zacatecas (Table A1), which is shown in Appendix A.

#### 2.1.2. Estimation Grid and Monitoring Zones

An estimation grid was defined in the study area, with nodes separated by 2 km in longitude and 2 km in latitude (291). The financial budget for monitoring was sufficient for three working days (8 h each). It was decided to divide the study area into three zones, investing one working day for each of them. A total of 84 nodes of the estimation grid corresponded to Zone C1, 140 to Zone C2, and 143 to Zone C3 (Figure 2).

The matrix of route distances between all the sites (including the base and all the monitoring wells) was built using the measurement tool incorporated in the digital map of the National Institute of Statistics and Geography (INEGI in Spanish), which includes the roads and paths that connect the different communities of Mexico [35].



Figure 2. Monitoring sites, estimation grid, and zones C1, C2 and C3 considered in the application.

#### 2.2. Geostatistical Analysis

Geostatistics is composed of techniques to obtain minimum variance estimate values at unsampled sites or times [36].

A classical geostatistical analysis consists of three steps: an exploratory analysis (identification of outliers and evaluation of the data distribution function), the structural analysis (calculation of the sample variogram and the adjustment of a valid variogram model), and the kriging estimation. For the proposed method, the exploratory and structural analyses of a geostatistical procedure were used to derive a variogram model and its corresponding covariance matrix, but the Kalman filter was selected for the optimization estimation method instead of kriging to reduce computational efforts.

To complete the exploratory analysis, the histogram of data must approximate a gauss bell. Within the structural analysis, a theoretical variogram model is fitted to the sample variogram.

To select the variogram model, a cross-validation is carried out using the procedure called leave one out, in which each data are individually removed, and the others are used to estimate or predict at that site or time. The entire geostatistical procedures were performed in the R free software [37].

Finally, the covariance matrix was constructed with the parameters of the variogram model [38].

To evaluate the selected variogram model, the mean squared error (*MSE*), Equation (1), and the standard mean squared error (*SMSE*), Equation (2), were used.

The *MSE* is the average value of the square differences between the measured and estimated values.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} \{ [Z(\mathbf{x}_i) - Z^*(\mathbf{x}_i)]^2 \}$$
(1)

The *SMSE* measures the consistency between the calculated variances and the estimate error variance ( $\sigma_i^2$ ). It is defined as the average of the square of the differences between the measured and estimated values over the estimate error variance, at the point of observation. A value close to one indicates that the model is adequate.

$$SMSE = \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{[Z(\mathbf{x}_i) - Z^*(\mathbf{x}_i)]^2}{\sigma_i^2} \right\}$$
(2)

where *n* is the number of observations,  $Z(\mathbf{x}_i)$  is the value of the property at point  $\mathbf{x}_i$ ,  $Z^*(\mathbf{x}_i)$  is the estimated value at point  $\mathbf{x}_i$ , and  $\sigma_i$  is the estimated standard deviation.

## 2.3. Kalman Filter

The Kalman filter (KF) is an algorithm based on a set of mathematical expressions to obtain unbiased linear recursive estimates of minimum variance for a system with noise. The recursive term refers to the fact that the filter recalculates the solution each time a new value or measurement is included [31]. The static Kalman filter version used in this paper is implemented with the following equations:

Measurement model: This links the measurement vector  $\mathbf{z}_n$ , which is presented in Equation (3), with the current state of the system  $C_n$  through the measurement matrix  $H_n$  and a Gaussian white noise  $v_n$  with covariance matrix  $R_n$ .

$$\boldsymbol{z_n} = \boldsymbol{H_n}\boldsymbol{C_n} + \boldsymbol{v_n} \tag{3}$$

where  $\boldsymbol{v}_n \sim N(0, R_n)$ .

In the update phase, the Kalman gain (Equation (4)) is calculated to obtain the new state vector  $\hat{C}_{n+1}^{n+1}$  (Equation (5)) and its covariance matrix  $P_{n+1}^{n+1}$  (Equation (6)):

$$\boldsymbol{K_{n+1}} = P_{n+1}^n H_{n+1}^T (H_{n+1} P_{n+1}^n H_{n+1}^T + \boldsymbol{R_{n+1}})$$
(4)

$$\widehat{\boldsymbol{C}}_{n+1}^{n+1} = \widehat{C}_{n+1}^n + K_{n+1} \Big( \boldsymbol{z}_{n+1} - H_{n+1} \widehat{\boldsymbol{C}}_{n+1}^n \Big)$$
(5)

$$P_{n+1}^{n+1} = (I - K_{n+1}H_{n+1})P_{n+1}^n$$
(6)

where the subscripts with n are the initial arrays and vectors, the subscripts n + 1 are the arrays and vectors resulting from the first calculation, and the superscripts are the arrays of the current operation.

# 2.4. Traveling Salesman Problem (TSP)

Ouaarab et al. [25] define the TSP as the shortest journey for a seller visiting different cities, adding the distance between them and visiting each one, before returning to the place of departure.

Following Cárdenas-Montes [26], the TSP (Equation (7)) can be expressed as:

$$x_{ij} = \begin{cases} 1 \text{ the route goes from city } i \text{ to city } j. \\ 0 \text{ another route.} \end{cases}$$
(7)

where  $x_{ij} = 1$  if city *i* is connected to city *j*;  $x_{ij} = 0$  otherwise.

For i = 0, ..., n, take  $c_{ij}$  as the distance from city i to city j. Then, TSP can be represented as shown in Equation (8):

$$\min \sum_{i=0}^{n} \sum_{\substack{j \neq i, j=0 \\ j \neq i, j = 0}}^{n} c_{ij} x_{ij}$$

$$0 \le x_{ij} \le 1 \qquad i, j = 0, ..., n$$

$$\sum_{\substack{i=0, i \neq j \\ n}}^{n} x_{ij} = 1 \qquad j = 0, ..., n$$

$$\sum_{\substack{j=0, j \neq i \\ j = 0, j \neq i}}^{n} x_{ij} = 1 \qquad i = 0, ..., n$$
(8)

The optimal monitoring route, obtained through the implementation of the TSP, is determined with the Branch and Bound method [39]. This method contributes to the design of monitoring networks where it is necessary to visit each monitoring network in situ for data acquisition.

# Branch and Bound

This algorithm generates subsets of solutions (branching), which are pruned or discarded if they do not meet the values of the limits established to delimit the problem, until the optimal solution is found [39].

This can be used to solve an optimization problem  $c^{0}(\mathbf{x})$  with an acceptable optimization domain <sup>0</sup>, as seen in Equation (9).

Minimize	$\mathcal{C}^{0}(\mathbf{X})$	
	$g_1^0(\mathbf{x}) \ge 0$ ,	
subjected to	$g_2^{\bar{0}}(\mathbf{x}) \ge 0$ ,	(9)
	$g_m^0(\mathbf{x}) \ge 0$	
and	$\in {}^{0}$	

where **x** represents a vector  $(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n)$ . The solution is said to be optimal if **x** satisfies all the constraints and belongs to  ${}^0$  and  $c^0(\mathbf{x})$  is the minimum. However, if the problem  $c^0(\mathbf{x})$  is very complicated, it can be broken down into simpler problems  $p_j$  (branching) whose optimal solutions are  $\mathbf{x}_j$ . It is necessary to include bounded problems  $p_k$ . The number of restrictions (*m*) changes according to the problem. This is represented as Equation (10), proposed by Lawler and Wood [40].

$$\begin{array}{c}
c^{j}(\mathbf{x}),\\
g_{1}^{j}(\mathbf{x}) \geq 0,\\
g_{2}^{j}(\mathbf{x}) \geq 0,\\
\dots\\g_{m}^{j}(\mathbf{x}) \geq 0\\
\in ,
\end{array}$$
(10)

# 2.5. Objective Function for the Design of the Monitoring Network

The proposed objective function (OF) represented in Equation (11) considers two variables and their respective weight, which can vary depending on the importance given to each variable. These variables are the priorities of the monitoring wells based on their contribution to reducing the total estimate error variance for the groundwater levels (obtained with the KF) and the optimal sampling route (solving the TSP).

$$OF = w_v P V_i + w_r P R_i \tag{11}$$

where  $w_v$  and  $w_r$  are the weights assigned to the variables  $PV_i$  (the priority assigned to well *i* based on the reduction in the total estimate variance it produces) and  $PR_i$  (priority given to well *i* based on the optimal distance to visit it and return to the starting point). The priority  $PV_i$  assigned to each well was obtained with the static Kalman filter, following the method proposed by Júnez-Ferreira [31]: a  $PV_i$  equal to one will be assigned to the well that produces the largest reduction in the total estimate error variance, while a  $N - n_{[MN]}$  value (N is the total number of wells available for the MN design and  $n_{[MN]}$  is the number of wells included in the optimal monitoring network (**MN**) will correspond to the last selected well). The TSP was used to obtain the optimal route to visit well *i* and returning to the starting point (the base). A  $PR_i$  equal to one will be assigned to the well for which the minimum distance is required, while a priority equal to  $N - n_{[MN]}$  will be assigned to the well with the longest route.

The first well selected for the monitoring network will correspond to the position i for which the lowest *OF* value was obtained. For those cases where the sum is the same for two positions, then the well that provides more information in terms of the total estimate error variance reduction (with a lower *PV*<sub>i</sub> value) is selected. Each time a new well is included in the network, the covariance matrix is updated, and the method is applied to test the remaining positions (at this stage, the TSP includes visiting the previously selected wells and the one that is being tested).

The sampling time considered at each visited site was 0.5 h. An average speed of 40 km/h was also taken to travel through the communication routes. The wells to be monitored in a predefined region will correspond to those where the necessary time for monitoring (the sum of sampling and travel times) is lower than or equal to 8 h (a working day). If there is additional budget to invest in another working day, the method was repeated, starting with the final covariance matrix of the previous day, and a new routing was initiated. In cases where all the wells of the MN have already been integrated to the **MN** and the working day has not finished, the method stops when a priority is assigned to each of them. The TSP is solved with the Branch and bound technique [39], using the network modeling (NET) version 1 module of the WinQSB software [41] through a Windows XP virtual machine [42]. The *OF* was evaluated using Microsoft Excel Version 2021 [43].

The proposed method is explained in the flow diagram of Figure 3. To fully understand this, it is necessary to declare the following elements: *i* is the number assigned to each well that is tested (i = 1, ..., N), n is the sum of the wells included in [MN] plus the *i* well with the minimum *OF*, rn is the distance of the optimal route to visit the wells that were already included in [**MN**] and the *i* well with the minimum *OF*, the vector [**MN**] includes the wells of the optimal monitoring network and [**MN**]<sup>*t*</sup> is the vector with the wells of the optimal monitoring network ordered following the optimal route determined with the TSP.



Figure 3. Diagram of the proposed method.

Process 1, included in Figure 3, consists of the following steps:

- 1. For each *i* well that is not yet part of the [**MN**], the TSP is applied (including the previously selected wells) to compute the optimal route (r<sub>i</sub>) from the starting point and back.
- 2. A  $PR_i$  is assigned to each well that is not part of the [**MN**],  $PR_i = 1$  for the *i* well with minimum r<sub>i</sub>,  $PR_i = N n_{[MN]}$  for the i well with maximum r<sub>i</sub>.
- 3. A  $PV_i$  is assigned to each well that is not part of the [**MN**] following the Júnez-Ferreira method [31];  $PV_i = 1$  for the *i* well that produces the largest reduction in the total estimate error variance, and  $PV_i = N n_{[MN]}$  for the *i* well that was last selected.

## 3. Results and Discussions

The method was applied to the existing monitoring network of the study area, which consists of 49 wells. Evidently, a frequent visit to all the wells of the existing monitoring network would be time-consuming and would lead to high costs. The proposed method was designed to reduce the number of wells selected in monitoring schemes without losing valuable information and reducing costs.

Although the optimization was performed for separate zones, the variogram was constructed with all the available groundwater-level data for a better representation of the spatial correlation for the groundwater levels in the study zone. Finally, a global model is preferred to estimate the groundwater levels within the aquifer to maximize the contribution of all the data collected at the selected monitoring sites (not only those located within a zone). Furthermore, the calculation of separate variograms for each zone could lead to a poor representation of the spatial correlation due to the limited number of data available for each of them [44].

The method was applied by assigning a  $w_v = 0.5$  and  $w_r = 0.5$ , which means that it was given the same importance to the reduction in the estimate error variance and the traveled distances.

As part of the geostatistical analysis, the available groundwater-level data were normalized. Statistics for these transformed data are shown in Table 1.

Statistics	Groundwater-Level	Standardized Groundwater-Level
Number of data	49	49
Minimum (m)	1988.44	-2.32
Maximum (m)	2272.53	2.32
Mean (m)	2081.9	0
Median (m)	2071.4	0
Standard deviation (m)	59.452	0.9974
Skewness	0.8647	0
Kurtosis	3.5552	2.7255

Table 1. Statistics of groundwater levels and standardized groundwater levels data.

Figure 4 shows the values of the adjusted variogram model and its parameters (nugget, sill and range) that allowed for the lowest value of the *MSE* and an acceptable *SMSE* value (close to one) to be obtained. The large value in the range reflects the strong correlation between the normalized data.

The sample variogram was computed for a lag size of 1800 m, a spherical variogram model was fitted to the sample variogram, and the model parameters (sill, nugget and range) were used to build the covariance matrix. A cross-validation (CV) (Figure 5) was used to evaluate the selected variogram model.



Figure 4. Sample variogram and adjusted model.

The value of the model range reflects that the correlation between the normalized data remains at large distances (almost reaching the extent of the estimation grid); however, the contribution of distant wells in the node estimation will be low, due to the screening effect.



Figure 5. Cross-validation results.

# Monitoring Network Optimization

Figures 6–8 show the final route to follow to monitor the  $n_{[MN]}$  wells included in the  $[MN]_{f}$  vector when applying the proposed method for zones C1, C2 and C3, separately (the label in these wells includes the well number and the route priority). Likewise, the

route segments to visit each well are presented in different colors. For each zone, the beginning and the end of the route is the base (UAZ campus).

### Monitoring Schemes for Zones C1, C2 and C3

According to the proposed method, seven wells of Zone C1 were selected to be visited within an estimated time of 7.78 h. The selected wells following the optimal route are: 47, 6, 2, 4, 48, 9 and 8, as shown in Figure 6.



**Figure 6.** Monitoring wells with their visit priorities for Zone C1, the monitoring route begins and ends at the UAZ (Base).

For Zone C2, it is possible to visit nine wells with a calculated time of 7.77 h. The sequence for visiting these selected wells was 43, 18, 32, 13, 11, 12, 17, 15 and 16, as presented in Figure 7.

When applying the proposed method for Zone C3, nine wells were selected to be visited in 6.92 h; visiting 10 wells leads to a routing of 8.07 h, exceeding the restriction of 8 h (a working day). However, since the extra time corresponds to only 4 min, the latter case was accepted. The chosen wells following the optimal monitoring route are 40, 31, 28, 29, 26, 27, 25, 24, 23 and 49, as illustrated in Figure 8.

For the same case study, the Júnez-Ferreira method was applied (considering only the estimate error variance reduction), which resulted in the monitoring scheme presented in Figure 9a–c for zones C1, C2 and C3, respectively.

Table 2 shows a comparison between the monitoring routes in the different zones, applying both methods. Following the proposed method, the travel distance was reduced by 28.95 km, 33.98 km and 18.96 km, in zones C1, C2 and C3, respectively, compared to the distances obtained with the Júnez-Ferreira method. An average reduction of 27.63 km was achieved; furthermore, more wells are visited when the proposed method is applied, which offers the advantage of collecting a higher number of field data with a lower cost.

Considering a vehicle performance of 6 km/L and a fuel cost of 0.99 USD per liter, visiting the 26 wells selected with the proposed method represents a total cost in transportation of 70.15 USD. This increases up to 83.66 USD for visiting the 21 wells selected without optimal routing. The total travel distance for visiting the 21 wells in three working



days was 507.07 km, whereas 425.18 km was necessary to visit the 26 wells selected with the proposal.

**Figure 7.** Monitoring wells with their visit priorities for Zone C2; the monitoring route begins and ends at the UAZ (Base).



**Figure 8.** Monitoring wells with their visit priorities for Zone C3; the monitoring route begins and ends at the UAZ (Base).



**Figure 9.** Routes and monitoring order of the wells selected when applying the Júnez-Ferreira method for zones C1 (**a**), C2 (**b**) and C3 (**c**).

Figure 10 shows the maps of groundwater-level estimates for the complete network (49 wells), the network obtained with the proposed method (26 wells), and the network obtained using the Júnez-Ferreira method (21 wells).

As expected, a more detailed description of the spatial configuration of groundwater levels within the study zone was obtained for an estimation using all the data collected with the complete network (Figure 10a). It also provides the lower estimated error variances. The estimation using the groundwater-level data for the wells selected with the proposed method (Figure 10b) reflects this flow pattern better than the estimates using the wells selected using the Júnez-Ferreira method (Figure 10c). Notably, the extended drawdown cone at the northeast area is not visible when the latter monitoring network is used. These differences are due to the different numbers and location of groundwaterlevel data used for each case; however, in the case of the northwest and southeast areas, the estimates produce similar spatial configurations for this variable.

Table 2. Comparison between the different route	s for the optimal	Calera aquifer	monitoring net-
work, selected with the proposed and the Júnez-Fe	erreira methods.		

Concept	ot Zone C1		Zone C2		Zone C3	
Mathada	Proposed	Júnez-Ferreira	Proposed	Júnez-Ferreira	Proposed	Júnez-Ferreira
Methods	Method	Method	Method	Method	Method	Method
Monitoring route	UAZ (Base)-Base C1-47-6-2-4-48-9- 8-Base C1-UAZ	UAZ (Base)-Base C1-6-38-4-1-46-9- Base C1-UAZ	UAZ (Base)-Base C2-43-18-32-13- 11-12-17-15-16-	<sup>2</sup> UAZ (Base)-Base- 43-14-11-39-15-16- 10-Base C2-UAZ	UAZ (Base)-40- 31-28-29-26-27- 25-24-23-49-	- UAZ (Base)-40- 28-29-26-49-33- 24-23-UAZ
	(Base)	(Base)	Base C2-UAZ (Base)	(Base)	UAZ (Base)	(Base)
Travel dis- tance (km)	171.27	200.22	131.1	165.08	122.81	141.77

Total time (h)	7.78	8.01	7.77	7.63	8.07	7.54
Number of wells visi- ted per wor- king day	7	6	9	7	10	8



**Figure 10.** Groundwater-level estimations considering complete network (**a**), with the proposed method (**b**), and with the Júnez-Ferreira method (**c**).

The estimate error variance maps for the complete network (49 wells), the network with the proposed method (26 wells), and the network with the Júnez-Ferreira method (21 wells), are displayed in Figure 11a–c. The variances obtained with the estimation using the wells for the three cases are very similar, which shows that the monitoring networks selected with the optimization methods are very effective.

The higher density of wells selected with the monitoring network obtained with the proposed method considerably reduces the estimation error variances in the valleys of the study zone (the middle strip that goes from south to north). These values are larger for the estimation using the data for the wells selected with the Júnez-Ferreira method. However, lower values are obtained around some wells located at peripheral areas of the estimation grid. These wells were discarded with the *OF* of the proposed method due to the increase in travel distance they produced.



**Figure 11.** Estimate error variances for groundwater levels considering: (**a**) the complete network, (**b**) the proposed method, and (**c**) the Júnez-Ferreira method.

Table 3 shows the statistics for the estimates using the selected monitoring networks for both approaches, and their comparisons with the estimates using the complete monitoring network. Taking the estimates using the complete monitoring network as a reference, the results show that the root of the total estimate error variance (the average standard error) increases up to 3.34 m when using the Júnez-Ferreira method and 3.35 m for the proposed method. The differences between both monitoring options for this value are negligible.

The mean difference, the mean square difference (MSD) and the root of the MSD values reflect the advantage of using a higher number of wells in the optimal monitoring network (26 vs. 21). Although, for the case of 26 wells, the maximum difference is larger than that obtained for the other monitoring option, the minimum and maximum values are more balanced.

Statistics	Complete Monitor ing Network (49 Wells)	Selected Monitor - ing Network with the Proposed Method (26 Wells)	Selected Monitor- ing Network with the Júnez-Ferreira Method (21 Wells)
Average standard error (m)	35.01	38.36	38.35
Mean difference (m)		5.76	12.46
Mean Square difference (m <sup>2</sup> )		179.76	394.68
Root of the Mean Square Difference (m)		13.41	19.87
Maximum difference (m)		74.92	86.63
Minimum difference (m)		-57.38	-19.83

Table 3. Statistics of the groundwater-level estimates in the Calera aquifer.

The maps of the differences between the groundwater-level estimates using the monitoring networks of 26 (proposed method) and 21 (neglecting optimal routes) wells, and the estimates obtained with the complete monitoring network of 49 wells are shown in Figure 12a,b, respectively. A larger area with low differences (between –10 and 10 m) was obtained for the monitoring network selected with the proposed method (26 wells).

For the monitoring network of 26 wells, the <-10 m differences cover a slightly larger area than that for the monitoring network of 21 wells. The extreme negative differences (<-19.83) that only exist for the monitoring network of 26 wells represent small portions of the southern, central and northwest. On the other hand, the areas for positive differences > 20 m are much larger when using the monitoring network of 21 wells. This result confirms the advantage of the proposed method in increasing the amount of collected data (five additional wells) to obtain a better spatial configuration of the groundwater levels.

These differences are associated with the coverage achieved by the wells selected for each monitoring scheme. Negative differences reflect that at least one well with a higher value than its neighbors was not included in the monitoring network (for example, well 1 is absent in Figure 12a). Conversely, positive values indicate that at least one well with a lower value than its neighbors was not included in the monitoring network (for example, well 31 in Figure 12b).





## 4. Conclusions

The proposed method allows for the design of groundwater-level monitoring networks based simultaneously on spatial estimation and the routes to follow for data collection; it considers the contribution of the available wells in the reduction in the total estimate error variance over a study zone, while looking for the minimum travel distance, thus reducing monitoring costs. This method could be adapted to the design of groundwater quality monitoring networks by simultaneously using the covariance matrices of the groundwater quality parameters considered within the optimization procedure.

The obtained results show that a monitoring network designed with the proposed method considerably reduces the monitoring costs (visiting 26 wells in three working days) without a significant loss of the amount of information that is obtained with respect to the reference monitoring network (49 wells). Neglecting the travel distances in the optimization procedure leads to the selection of only 21 wells to be monitored over three working days, which implies that the proposal allows more field data to be obtained at the same cost. Furthermore, the higher number of sampling wells selected with the proposal produces a spatial configuration of groundwater levels that are more similar to those obtained with the reference monitoring network.

An important feature of the proposed method is its additive nature, which means that new wells can be incorporated into the monitoring network if the budget increases without eliminating the previously selected monitoring sites, favoring the continuity of the groundwater-level time series that is collected with them.

For the case study, it is assumed that the routes to follow correspond to the same type of communication route, in which a constant speed of the vehicle can be maintained during travel. Additionally, the same sampling time was considered for each visited well. Future works to improve the method could integrate specific aspects that produce variations in monitoring conditions, such as traveling on dirt roads, highways, ease of measuring, well depth, dead times, lunch time, etc. Another interesting option is the inclusion of the temporal component in the objective function, using time series information.

The monitoring network selected with the proposal very slightly increases the total estimate error variance compared to the case when the travel distances are neglected during the optimization, but produces estimates that are closer to the complete monitoring

network. Additionally, the efficiency in the travel distance allows more wells to be visited in a working day, which implies a greater amount of direct (field) information at the same cost.

Future work should be oriented toward the development of a user-friendly software that integrates the different routines that are necessary for the optimization of the monitoring network, including a procedure to automatically generate the route matrix from a map with the different roads connecting the tested monitoring sites.

Author Contributions: Conceptualization: J.C.E., H.E.J.-F. and C.O.R.R.; methodology, validation, and writing: J.C.E. and H.E.J.-F.; resources and supervision: J.G.-T., C.B.-C. and C.O.R.R.; investigation: J.C.E.; writing—original draft: J.C.E.; writing—review and editing: J.C.E., H.E.J.-F., J.G.-T., C.B.-C. and C.O.R.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available in the article.

**Acknowledgments:** The authors express their gratitude to the Mexican National Council for Science and Technology (CONACYT) for financing the scholarship of the doctoral student. We deeply appreciate the recommendation and professional comments from the reviewers and the editor.

Conflicts of Interest: The authors declare no conflict of interest.

# Appendix A

**Table A1.** Location of the monitoring network and groundwater levels for the Calera aquifer (2017).

Wall Number	Wall Nama			Water Level
wen number	wen name	UTWIX		(masl)
1	N-5	722,350.43	2,566,447.24	2193.86
2	N-6	727,635.07	2,569,112.14	2071.40
3	N-6′	729,084.62	2,567,334.11	2049.09
4	S/N	734,491.45	2,567,581.45	2046.78
5	N-9	726,766.72	2,565,433.71	2082.56
6	N-11	726,794.37	2,564,738.64	2179.70
7	N-12	733,473.68	2,561,912.06	2015.18
8	N-13	740,173.85	2,561,855.34	2042.99
9	N-14	745,805.00	2,562,445.68	2061.46
10	N-16	725,781.34	2,556,285.20	2045.21
11	N-19	741,341.57	2,557,223.85	2033.42
12	N-22	728,225.67	2,553,180.33	2021.23
13	N-23	740,152.13	2,551,689.60	2018.38
14	N-24	745,389.30	2,550,835.77	2014.00
15	N-25	725,838.04	2,546,666.41	2150.80
16	N-26	732,058.87	2,544,305.51	2146.40
17	N-27	736,679.40	2546,528.74	2046.46
18	N-28	741,053.95	2,546,610.64	2058.50
19	N-30 A	746,475.09	2,548,025.16	2034.61
20	N-30 C	734,125.96	2,543,297.39	2120.39
21	N-31	734,994.10	2,541,532.17	2103.91
22	N-33	739,935.57	2,542,607.28	2089.70
23	SUST.	725,659.20	2,536,936.50	2088.52
24	N-35′	735,102.00	2,534,064.90	2100.25
25	N-37	737,337.29	2,536,060.23	2103.44
26	S/N	735,770.90	2,529,933.11	2183.58

27	S/N	739,406.11	2,533,240.06	2099.34
28	N-41	738,287.73	2,524,768.56	2155.56
29	N-42	732,387.85	2,520,864.15	2141.80
30	S/N	739,167.82	2,521,061.75	2135.57
31	C.P 4	734,931.70	2,523,728.33	2088.81
32	SN RAFAEL	741,669.08	2,551,221.58	2010.96
33	S/N	741,082.87	2,537,692.27	2080.30
34	S/N	743,570.81	2,548,288.58	2039.11
35	S/N	743,557.28	2,554,902.28	2037.05
36	TALANCON	739,107.71	2,559,477.68	2037.95
37	N-15	741,398.13	2,527,950.69	2138.17
38	VIVERO	736,282.30	2,570,173.41	2030.83
39	S/N	736,794.51	2,548,441.69	2070.49
40	S/N	739,850.11	2,518,333.57	2128.86
41	S/N	734,796.27	2,527,680.68	2073.10
42	S/N	737,430.66	2,519,920.46	2160.77
43	S/N	738,393.26	2,541,570.25	2113.20
44	S/N	738,409.40	2,549,529.02	2046.07
45	S/N	736,172.37	2,559,163.21	1988.44
46	S/N	734,674.42	2,560,296.74	2011.40
47	S/N	736,874.70	2,562,436.55	2011.59
48	S/N	731,673.45	2,564,949.14	2038.76
49	JESUS LOPEZ	744,739.46	2,529,823.22	2272.53

# References

- 1. CONAGUA-AAM. Available online: http://sina.conagua.gob.mx/publicaciones/AAM\_2018.pdf (accessed on 24 August 2021).
- 2. Food and Agriculture Organization (FAO). Available online: https://www.fao.org/aquastat/es/overview/methodology/water-use (accessed on 8 September 2021).
- 3. CONAGUA-EAM. Available online: http://sina.conagua.gob.mx/publicaciones/EAM\_2018.pdf (accessed on 23 August 2021).
- 4. Bhat, S.; Motz, L.H.; Pathak, C.; Kuebler, L. Geostatistics-based groundwater-level monitoring network design and its application to the Upper Floridan aquifer, USA. *Environ. Monit. Assess.* **2015**, *187*, 4183.
- 5. Molinari, A.; Guadagnini, L.; Marcaccio, M.; Guadagnini, A. Geostatistical multimodel approach for the assessment of the spatial distribution of natural background concentrations in large-scale groundwater bodies. *Water Res.* **2019**, *149*, 522–532.
- Ahmadi, S.H.; Sedghamiz, A. Geostatistical Analysis of Spatial and Temporal Variations of Groundwater Level. *Environ. Monit.* Assess. 2007, 129, 277–294.
- 7. International Groundwater Resources Assessment Centre (IGRAC). Available online: https://www.un-igrac.org/sites/de-fault/files/resources/files/GGMN%20Brochure%202016.pdf (accessed on 3 November 2021).
- Briseño-Ruiz, J.-V.; Júnez-Ferreira, H.-E.; Herrera-Zamarrón, G.-S. Method for the optimal design of networks to monitor groundwater levels. Water Technol. Sci. 2011, 2, 77–96.
- 9. Kumari, K.; Jain, S.; Dhar, A. Computationally efficient approach for identification of fuzzy dynamic groundwater sampling network. *Environ. Monit. Assess.* **2019**, *191*, 310.
- Mirzaie-Nodoushan, F.; Bozorg-Haddad, O.; Loáiciga, H.A. Optimal design of groundwater-level monitoring networks. J. Hydroinform. 2017, 19, 920–929.
- 11. Soltani, S.; Kordestani, M.; Karim Aghaee, P. New estimation methodologies for well logging problems via a combination of fuzzy Kalman filter and different smoothers. *J. Pet. Sci. Eng.* **2016**, *145*, 704–710.
- 12. Bierkens, M.F.; Knotters, M.; Hoogland, T. Space-time modeling of water table depth using a regionalized time series model and the Kalman filter. *Water Res.* 2001, *37*, 1277–1290.
- 13. Luo, Q.; Wu, J.; Yang, Y.; Qian, J.; Wu, J. Multi-objective optimization of long-term groundwater monitoring network design using a probabilistic Pareto genetic algorithm under uncertainty. *J. Hydrol.* **2016**, *534*, 352–363.
- 14. Farlin, J.; Gallé, T.; Pittois, D.; Bayerle, M.; Schaul, T. Groundwater quality monitoring network design and optimisation based on measured contaminant concentration and taking solute transit time into account. *J. Hydrol.* **2019**, *573*, 516–523.
- Song, J.; Yang, Y.; Chen, G.; Sun, X.; Lin, J.; Wu, J.; Wu, J. Surrogate assisted multi-objective robust optimization for groundwater monitoring network design. J. Hydrol. 2019, 577, 123994.
- 16. Azadi, S.; Amiri, H.; Ataei, P.; Javadpour, S. Optimal design of groundwater monitoring networks using gamma test theory. *Hydrogeol. J.* **2020**, *28*, 1389–1402.

- Elshall, A.S.; Ye, M.; Finkel, M. Evaluating two multi-model simulation-optimization approaches for managing groundwater contaminant plumes. J. Hydrol. 2020, 590, 125427.
- Ondrasek, G.; Bakić Begić, H.; Romić, D.; Brkić, Ž.; Husnjak, S.; Bubalo Kovačić, M. A novel LUMNAqSoP approach for prioritising groundwater monitoring stations for implementation of the Nitrates Directive. *Environ. Sci. Eur.* 2021, 33, 23.
- 19. Nourani, V.; Ejlali, R.G.; Alami, M.T. Spatiotemporal Groundwater Level Forecasting in Coastal Aquifers by Hybrid Artificial Neural Network-Geostatistics Model: A Case Study. *Environ. Eng. Sci.* **2011**, *28*, 217–228.
- Uddameri, V.; Kakarlapudi, C.; Hernandez, E.A. A GIS enabled nested simulation-optimization model for routing groundwater to overcome spatio-temporal water supply and demand disconnects in South Texas. *Environ. Earth Sci.* 2014, 71, 2573–2587.
- Manzione, R.L.; Wendland, E.; Tanikawa, D.H. Stochastic simulation of time-series models combined with geostatistics to predict water-table scenarios in a Guarani Aquifer System outcrop area, Brazil. *Hydrogeol. J.* 2012, 20, 1239–1249.
- 22. 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.* **2013**, *185*, 1–19.
- Zhou, Y.; Dong, D.; Liu, J.; Li, W. Upgrading a regional groundwater level monitoring network for Beijing Plain, China. *Geosci.* Front. 2013, 4, 127–138.
- 24. Ran, Y.; Li, X.; Ge, Y.; Lu, X.; Lian, Y. Optimal selection of groundwater-level monitoring sites in the Zhangye Basin, Northwest China. J. Hydrol. 2015, 525, 209–215.
- Ouaarab, A.; Ahiod, B.; Yang, X.-S. Random-key cuckoo search for the travelling salesman problem. Soft Comput. 2015, 19, 1099– 1106.
- 26. Cardenas-Montes, M. Creating hard-to-solve instances of travelling salesman problem. Appl. Soft Comput. 2018, 71, 268–276.
- 27. Hatamlou, A. Solving travelling salesman problem using black hole algorithm. *Soft Comput.* **2018**, *22*, 8167–8175.
- 28. Miranda, P.-A.; Blazquez, C.A.; Obreque, C.; Maturana-Ross, J.; Gutierrez-Jarpa, G. The bi-objective insular traveling salesman problem with maritime and ground transportation costs. *Eur. J. Oper. Res.* **2018**, *271*, 1014–1036.
- 29. Hu, Y.; Yao, Y.; Lee, W.-S. A reinforcement learning approach for optimizing multiple traveling salesman problems over graphs. *Knowl. Based Syst.* **2020**, *204*, 106244.
- Nunes, L.M.; Cunha, M.C.; Ribeiro, L. Optimal Space-time Coverage and Exploration Costs in Groundwater Monitoring Networks. *Environ. Monit. Assess.* 2004, 93, 103–124.
- 31. Júnez-Ferreira, H.-E. Diseño de una Red de Monitoreo de la Calidad del Agua Para el Acuífero Irapuato-Valle, Guanajuato. Master's Thesis, Universidad Nacional Autónoma de México, Ciudad de México, Mexico, 2005.
- 32. CONAGUA. Disponibilidad Media Annual de Agua En El Acuífero de Calera Estado de Zacatecas. Available online: https://sigagis.conagua.gob.mx/gas1/Edos\_Acuíferos\_18/zacatecas/DR\_3225.pdf (accessed on 11 September 2021).
- INEGI. Available online: http://cuentame.inegi.org.mx/monografias/informacion/zac/territorio/div\_municipal.aspx?tema=me&e=32 (accessed on 10 January 2022).
- Flores-Rodarte, A.; Cristobal-Acevedo, D.; Pascual-Ramírez, F.; León-Mojarro, B.; Prado-Hernández, J.-V. Agricultural water productivity in the central zone of the Calera aquifer, Zacatecas. Agric. Eng. Biosyst. 2019, 11, 181–199.
- 35. INEGI 2022. Digital Map of Mexico V6. Available online: http://gaia.inegi.org.mx/mdm6/ (accessed on 15 February 2022).
- 36. Georgakakos, A.-P.; Kitanidis, P.-K.; Loaiciga, H.-A.; Olea, R.-A.; Yates, S.-R.; Rouhani, S. Review of geostatistics in geohidrology I: Basic concepts. J. Hydraul. Eng. 1990, 116, 612–632.
- RStudio, Inc. Version 1.1.463–© 2022–2018. Available online: https://www.rstudio.com/products/rstudio/older-versions/ (accessed on 23 March 2022).
- Oliver, M.-A.; Webster, R. A tutorial guide to geostatistics: Computing and modelling variograms and kriging. *Catena* 2014, 113, 56–69.
- Narendra, P.M.; Fukunaga, K. A branch and bound algorithm for feature subset selection. *IEEE Trans. Comput.* 1977, 26, 917–922.
- 40. Lawler, E.-L.; Wood, D.-E. Branch-and-bound methods: A survey. Oper. Res. 1966, 14, 699–719.
- 41. WinQSB 2.0. Network Modeling Version 1. Available online: https://winqsb.uptodown.com/windows (accessed on 12 February 2022).
- 42. VirtualBox 6.1, Version 6.1.14 r140239 (Qt5.6.2). Available online: https://www.virtualbox.org/ (accessed on 21 January 2022).
- Microsoft Corporation. 2022. Microsoft Excel Version 2021. Available online: https://office.microsoft.com/excel (accessed on 23 February 2022).
- 44. Júnez-Ferreira, H.-E.; Herrera, G.-S., Saucedo, E.; Pacheco-Guerrero, A. Influence of available data on the geostatistical-based design of optimal spatiotemporal groundwater-level-monitoring networks. *Hydrogeol. J.* 2019, 27, 1207–1227.