Environ Monit Assess (2016) 188: 39
DOI 10.1007/s10661-015-5036-y

Optimal design of monitoring networks for multiple groundwater quality parameters using a Kalman filter: application to the *Irapuato-Valle* aquifer

H. E. Júnez-Ferreira · G. S. Herrera · L. González-Hita · A. Cardona · J. Mora-Rodríguez

Received: 3 June 2014 / Accepted: 8 December 2015 / Published online: 17 December 2015

Abstract A new method for the optimal design of groundwater quality monitoring networks is introduced in this paper. Various indicator parameters were considered simultaneously and tested for the Irapuato-Valle aquifer in Mexico. The steps followed in the design were (1) establishment of the monitoring network objectives, (2) definition of a groundwater quality conceptual model for the study area, (3) selection of the parameters to be sampled, and (4) selection of a monitoring network by choosing the well positions that minimize the estimate error variance of the selected indicator parameters. Equal weight for each parameter was given to most of the aquifer positions and a higher weight to priority zones. The objective for the monitoring network in the specific application was to obtain a general reconnaissance of the water quality, including water types, water origin, and first indications of contamination. Water quality indicator parameters were chosen in accordance with this objective, and for the selection of the optimal monitoring sites, it was sought to obtain a low-uncertainty estimate of these parameters for the entire aquifer and with more certainty in priority zones. The optimal monitoring network was selected using a combination of geostatistical methods, a Kalman filter and a heuristic optimization method. Results show that when monitoring the 69 locations with higher priority order (the optimal monitoring network), the joint average standard error in the study area for all the groundwater quality parameters was approximately 90 % of the obtained with the 140 available sampling locations (the set of pilot wells). This demonstrates that an optimal design can help to reduce monitoring costs, by avoiding redundancy in data acquisition.

Keywords Optimal monitoring network · Geostatistics · Kalman filter · Groundwater quality · Priority zones

Introduction

In this paper, we introduce a Kalman filter (KF) methodology for the optimal design of groundwater quality

H. E. Júnez-Ferreira (✉)
Maestría en Ingeniería Aplicada, Universidad Autónoma de Zacatecas, Av. Ramón López Velarde No. 801, Carretera a la Bufa, C.P. 98010 Zacatecas, Mexico
e-mail: hejunez@hotmail.com

G. S. Herrera
Instituto de Geofísica, Universidad Nacional Autónoma de México, Ciudad Universitaria, Del. Coyoacán, CP 04510 Mexico City, Mexico

L. González-Hita
Subcoordinación de Hidrología Subterránea, Instituto Mexicano de Tecnología del Agua, Paseo Cuauhnáhuac 8532, Col. Progreso, C.P. 62550 Jiutepec, Morocco, Mexico

A. Cardona
Área Ciencias de la Tierra, Facultad de Ingeniería, Universidad Autónoma de San Luis Potosí, Dr. Manuel Nava No 8, Zona Universitaria, C.P. 78290 San Luis Potosí, S.L.P., Mexico

J. Mora-Rodríguez
Departamento de Ingeniería Geomática e Hidráulica, Universidad de Guanajuato, Av. Juárez, 77. Col. Centro, C.P. 36000 Guanajuato, Mexico

monitoring networks, in which various indicator parameters and priority zones are considered simultaneously. We test it in the design of a sampling network for the *Irapuato-Valle* aquifer in Mexico.

Background

Groundwater monitoring network design consists in choosing observation well positions and whenever it is possible also the monitoring frequencies to achieve predetermined monitoring objectives. When an optimal design is required, optimization criteria are established and an optimization method is used to choose sampling well positions and/or sampling frequencies that minimize or maximize it. A spatial design is shown in this paper by selecting only sampling positions.

In general, monitoring objectives involve estimating parameters at unobserved positions or times, and therefore, interpolation or estimation methods are frequently used in the design.

Optimal groundwater monitoring network design

There are several published papers for the optimal design of groundwater quality or groundwater quantity monitoring networks. Early research focused on methods to locate new monitoring wells. Afterward, methods were developed to identify sampling plans to minimize the spatial and/or temporal redundancy in existing monitoring networks (ASCE 2003).

Herrera and Pinder (2005) defined three main approaches for the design of groundwater monitoring networks (selection of positions and its monitoring program) as (1) hydrological, based on site hydrological conditions only; (2) statistical, based on inferences obtained from statistical analysis of data; and (3) modeling, based on results of groundwater flow and/or transport models.

The literature review shown below focuses on the statistical framework because the methodology presented in this paper is of the same kind. Methods for the design of monitoring networks for groundwater quality and quantity, relevant for our research, are included.

In the selection of positions and frequencies, most research use methods that optimize a function of the estimate error variance for a specific area. Within the statistical approach, most research are based on geostatistical techniques that consider spatial correlations between groundwater data. Some recent examples for water quality monitoring designs are Chadalavada et al. (2011), Li et al. (2011), and Hergt (2009). On the other hand, some examples for groundwater quantity monitoring designs are Kumar et al. (2005) and Zaidi et al. (2007). All these papers focus on the optimal design of monitoring networks for a single variable (e.g., the concentration of a solute in groundwater or the water level) in a spatial context where only positions were selected. Lin and Rouhani (2001) designed different groundwater quality monitoring networks for the two analyzed contaminants, trichloroethylene and tetrachloroethylene.

Only a few reported research have tried to incorporate various water quality parameters during optimization of a single monitoring network design. Masoumi and Kerachian (2010) claimed that the transinformation-based methodology they presented has the ability to consider various variables at the same time, but it was not demonstrated in the paper. Dutta et al. (1998) employed various water quality parameters to suggest different monitoring alternatives, but in the optimization process, only the parameter with larger variance was considered. Preziosi et al. (2012) applied map algebra and ranking score in a geographic information system (GIS) procedure to integrate aquifer vulnerability and levels of groundwater pollution in a monitoring network design, considering various water quality parameters.

Yeh et al. (2006) developed a methodology that is relevant for the problem we address. They employed a multivariate analysis to design a spatial monitoring network for nine water quality parameters. Variables were standardized before the analysis. A coregionalization matrix of the groundwater quality parameters was calculated using the direct and cross variograms that incorporate two structures to consider short and long spatial scale variations. The eigenvalues and the variance proportion for regionalized factors were calculated through the principal component analysis (PCA) and factorial analysis of the coregionalization matrix. An optimization problem that has three key characteristics was posed, which makes it different from the methods previously reviewed. First, the objective function seeks to minimize the estimate error variance of regionalized factors (obtained using factorial kriging) instead of using the estimate error variance of original variables. Second, the optimal monitoring network design can be carried out by considering different combinations of regionalized factors regardless of spatial scales. Third, the weighting for each regionalized factor (RF) during the optimization is assigned according to the variance proportion that represents. This differs from the kriging and cokriging-based methods, which use a subjective weight of regional variables. A genetic algorithm is used to get the optimal design.

The methodology introduced in this paper is based on the method of Herrera (1998) that was modified for its use in the design of an optimal monitoring network (OMN) for various water quality parameters. A function of the estimate error variance was employed as the criterion to choose the sampling wells of a monitoring network. Unlike Herrera (1998) that derives the elements of a covariance matrix for the considered water quality parameter from a numerical transport model, a covariance matrix was calculated for each water quality parameter from a geostatistical analysis in this paper. Furthermore, we consider jointly the spatial correlations of various groundwater quality parameters and its corresponding priority zones (that will be described in the Materials and methods section) in an automated procedure that uses a heuristic optimization method in combination with a Kalman filter to minimize a joint-normalized variance of all the parameters. Initial developments of the methodology introduced in this paper were presented in Herrera et al. (2004) and Júnez (2005).

Materials and methods

The steps followed in the OMN design were (1) establishment of the monitoring network objectives, (2) definition of a groundwater quality conceptual model for the study area, (3) selection of the parameters to be sampled, and (4) design of an OMN by choosing the well positions that minimize the estimate error variance of the selected indicator parameters. Equal weight was assigned to all the aquifer area except to priority zones for which a higher weight was used.

The design of a monitoring network begins by setting its objectives. These objectives depend on the needs of the water resource management organizations. The criteria used in the design of the monitoring network for the *Irapuato-Valle* aquifer will be explained later.

Field (geologic and hydrogeologic) and laboratory evidence were used to identify anthropogenic and geogenic contamination sources restraining groundwater quality for drinking purposes. Diffuse contamination sources from irrigation using raw wastewater increase nutrients (nitrate and phosphate) and microorganism concentrations in groundwater. Groundwater interaction along flow path with geogenic sources represented by igneous rocks (both felsic and mafic extrusive lava flows, tuffs, and ignimbrites) produce fluoride, arsenic, iron, and manganese concentrations above drinking water standards. This information was useful to establish the groundwater quality conceptual model and for the selection of key parameters to delineate the aquifer priority zones.

As mentioned before, the introduced method to select an OMN considers the spatial correlations of various groundwater quality parameters and its priority zones in an automated optimization procedure. A Kalman filter and an optimization method are used to choose the monitoring positions.

The Kalman filter requires a prior spatial covariance matrix that is calculated through geostatistics. The optimization method is heuristic and selects the spatial location that minimizes an objective function one at a time. In this paper, the optimization procedure seeks to reduce the estimate error variance over an area of interest (it could consider contaminated areas, highly vulnerable zones, pollution sources, potable water exploitation zones, recharge zones, or the entire area that defines an aquifer). To achieve this, a spatial set of locations is necessary, for which estimates are needed.

The Kalman filter

The KF is a set of mathematical equations that provide a minimum-variance unbiased linear estimate for the state of a system given noisy data (Jazwinski 1970). In this paper, we apply the static Kalman filter formulas presented for spatiotemporal monitoring designs in Júnez-Ferreira and Herrera (2013) in space for each analyzed water quality indicator parameter (WQIP) because the available information for the *Irapuato-Valle* aquifer was from a sampling campaign. Some examples of space-time applications can be found in Herrera (1998), Herrera and Pinder (2005), and Júnez-Ferreira and Herrera (2013).

The linear measurement equation of the discrete Kalman filter, which relates the state vector of the variable $\mathbf{q}$ in the positions of interest, with sampled data $\mathbf{z}$ is

$$\mathbf{z}_j = \mathbf{H}_j\mathbf{q} + \mathbf{v}_j \qquad (1)$$

where $\{\mathbf{z}_j, j = 1, 2, \ldots\}$ is a sequence of water quality measurements for a single parameter. The *j*th sampling matrix, $\mathbf{H}_j$, is a $1\times N$ matrix that is nonzero only at the position corresponding to the entry of $\mathbf{q}$ from where the jth sample is taken and N is the dimension of the vector $\mathbf{q}$. $\mathbf{q}=\{q_i\}$ is the spatial vector with the water quality parameter in the positions of interest (q_i is the water quality parameter in position x_i). The vector $\{\mathbf{v}_j, j=1,2,\ldots\}$ represents measurement errors; they are a white Gaussian sequence, with zero mean and covariance $\mathbf{r}_j$.

The measurement error sequence $\{\mathbf{v}_j\}$ and the vector $\mathbf{q}$ are independent. In the case study, the measurement error variance included in the Kalman filter formulations was very small because the measurement error was considered negligible.

The estimate error covariance matrix is

$$\mathbf{P}^n = E\left\{\left(\mathbf{q}-\hat{\mathbf{q}}^n\right)\left(\mathbf{q}-\hat{\mathbf{q}}^n\right)^T \Big/ \mathbf{z}_1,\mathbf{z}_2,\ldots,\mathbf{z}_n\right\} \tag{2}$$

where $\hat{\mathbf{q}}^n = E\ \{\mathbf{q}/\mathbf{z}_1,\mathbf{z}_2,\ldots,\mathbf{z}_n\}$, being $E\{\bullet/\mathbf{z}_1,\mathbf{z}_2,\ldots,\mathbf{z}_n\}$ the expected value of •, given the measurements $\mathbf{z}_1,\mathbf{z}_2,\ldots,\mathbf{z}_n$.

To implement the filter, prior estimates of $\mathbf{q}$ (named $\hat{\mathbf{q}}^0$) and of the error covariance matrix ($\mathbf{P}^0$) are required. Given these prior estimates, the minimum-variance linear estimate for $\mathbf{q}$ can be obtained sequentially through the following formulas:

$$\hat{\mathbf{q}}^{n+1} = \hat{\mathbf{q}}^n + \mathbf{K}_{n+1}\left(\mathbf{z}_{n+1}-\mathbf{H}_{n+1}\hat{\mathbf{q}}^n\right) \tag{3}$$

$$\mathbf{P}^{n+1} = \mathbf{P}^n-\mathbf{K}_{n+1}\mathbf{H}_{n+1}\mathbf{P}^n \tag{4}$$

$$\mathbf{K}_{n+1} = \mathbf{P}^n\mathbf{H}_{n+1}^T\left(\mathbf{H}_{n+1}\mathbf{P}^n\mathbf{H}_{n+1}^T + \mathbf{r}_{n+1}\right) \tag{5}$$

There are many ways to estimate $\hat{\mathbf{q}}^0$; a particular one is presented in the example included in this paper. The prior estimate error covariance matrix $\mathbf{P}^0$ for each WQIP is obtained through a geostatistical analysis of data measured from a field campaign, by fitting a spatial variogram model to the sample variogram. Once the variogram model is selected, the elements of the spatial covariance matrix are calculated with Eq. (6).

$$C(h) = C(0)-\gamma(h) \tag{6}$$

where $C(0)$ is the variance of the analyzed parameter, and it is equal to the sill of the variogram, and $\gamma(h)$ is the variogram model function. Equation (6) assumes that the variogram is bounded.

Monitoring network optimization

A groundwater quality monitoring network that considers simultaneously the uncertainty reduction for various parameters could help avoid spatial redundant information for all the parameters at once. The optimization procedure employed in this paper to consider various WQIPs simultaneously is explained below.

Consider the set of all possible monitoring positions $M=\{\mathbf{x}_i^M, i=1,\ldots,Nmp\}$, where Nmp is the number of possible monitoring positions, and $\mathbf{x}=(x,y)\in D$, where D is the Euclidean plane. From this set, we want to choose those points that minimize the sum of the estimate error variance on the points for which estimates are needed, $E=\{\mathbf{x}_j^E, j=1,\ldots,Nep\}$, where Nep is the number of estimation points.

In the specific application addressed in the paper, the statistical objective for the monitoring network was to obtain low-uncertainty estimates of groundwater quality for the entire aquifer and with more certainty in priority zones for the most representative parameters. For the optimization procedure, this objective was "translated" in discrete mathematical terms. Because of this, a preliminary spatial estimation grid with squared elements of 2 km side length was defined and an auxiliary grid (composed by squared elements of 1 km side) was defined for each WQIP that encompasses its priority zone. The optimization objective is to choose the monitoring positions $\mathbf{x}_i^M\in M$ that minimize the joint total (JT) normalized variance of the estimation error, calculated using the following formula:

$$\sigma^2_{\mathrm{JT}} = \sum_{k=1}^{\mathrm{NWQIP}}\sum_{j=1}^{\mathrm{NP}}\sigma^2_{k,j} + \sum_{j=1}^{\mathrm{NPZ}_1}\sigma^2_{1,j} + \sum_{j=1}^{\mathrm{NPZ}_2}\sigma^2_{2,j} + \ldots + \sum_{j=1}^{\mathrm{NPZ}_{\mathrm{NWQIP}}}\sigma^2_{\mathrm{NWQIP},j} \tag{7}$$

where $\sigma^2_{k,j}$ is the normalized variance of the estimation error $e_k\left(\mathbf{x}_j^E\right) = q_k\left(\mathbf{x}_j^E\right)-\hat{q}_k\left(\mathbf{x}_j^E\right)$, of the parameter k at the jth estimation location, obtained from the diagonal of the KF covariance matrix of each parameter; NP is the number of elements of the preliminary spatial estimation grid; NWQIP is the number of water quality indicator parameters considered simultaneously in the design; and NPZ_i is the number of elements in the mesh of the priority zone of the ith WQIP. If we assign $\sigma^2_{k,j}=0$ at all the estimation positions of the auxiliary grids inside priority zones not corresponding to parameter k, then we can write

$$\sigma^2_{\mathrm{JT}} = \sum_{k=1}^{\mathrm{NWQIP}} \sum_{j=1}^{Nep} \sigma^2_{k,j} \tag{8}$$

where $N_{ep} = \mathrm{NP} + \sum_{i=1}^{\mathrm{NWQIP}} \mathrm{NPZ}_i$

It can be seen that if the variance of each parameter is used in Eqs. (7) and (8), more weight during the optimization procedure would be given to parameters with large variance than to parameters with small variance; therefore, low-variance parameters would be poorly monitored with the resulting monitoring network. In order to give the same weight to all the WQIPs in the optimization procedure, we normalized the covariance matrix obtained from the geostatistical analysis for each parameter to obtain a correlation matrix (CM). The normalization of each covariance matrix is carried out dividing all its elements by the parameter variance (the sill of its corresponding variogram model). In this way, each component of the principal diagonal for each prior CM is equal to one. This is,

$$\mathbf{CM}^0_k = \frac{\mathbf{C}^0_k}{C(0)_k} \tag{9}$$

where $\mathbf{CM}^0_k$ is the correlation matrix for parameter k, $\mathbf{C}^0_k$ is the covariance matrix obtained from the geostatistical analysis for parameter k, and $C(0)_k$ is the sill of the variogram model for parameter k.

The problem is optimized sequentially using a successive inclusion method (Samper and Carrera 1990), as explained below. For $n=1, \cdots$, choose the point $\mathbf{X}_{o,n}$ in M that minimizes the function

$$\sigma^2_{\mathrm{JT}}(\mathrm{OMN}(n-1), \mathbf{x}) = \sum_{k=1}^{\mathrm{NWQIP}} \sum_{j=1}^{Nep} \sigma^2_{k,j}(\mathrm{OMN}(n-1), \mathbf{x}) \tag{10}$$

where $\sigma^2_{k,j}(\mathrm{OMN}(n-1), \mathbf{x})$ is the element in the diagonal of the updated correlation matrix of parameter k corresponding to the estimation point $(\mathbf{x}^E_j)$ obtained from the KF using the $n-1$ optimal monitoring network $\mathrm{OMN}(n-1)=\{\mathbf{X}_{o,1}, \mathbf{X}_{o,2}, \ldots, \mathbf{X}_{o,n-1}\}$ previously chosen and the spatial point $\mathbf{x}$. It is important to note that OMN (0) is an empty set. Let $\mathbf{P}^{n-1}_{k,o}$ be the matrix of parameter k obtained after applying Eqs. (4) and (5) of the KF to the set OMN $(n-1)$. Then, to find $\mathbf{X}_{o,n}$, for each possible monitoring point that has not been chosen, the updated correlation matrix for each parameter is calculated using Eqs. (4) and (5) of the KF with the matrix $\mathbf{P}^{n-1}_{k,o}$ and the possible monitoring point, and then, the point that gives the minimum $\sigma^2_{\mathrm{JT}}(\mathrm{OMN}(n-1), \mathbf{x})$ is selected. This point is then added to the set of optimal monitoring points, and the matrix $\mathbf{P}^n_{k,o}$ for each parameter is calculated again using the KF. The process starts by assigning $\mathbf{P}^0_{k,o}=\mathbf{CM}^0_k$.

A flowchart of the proposed methodology is shown in Fig. 1. The FORTRAN program GWQMonitor_Geoestad that applies this algorithm, designed by the authors (Júnez 2005), was used to get the results.

Samper and Carrera (1990) showed that successive inclusion is a suboptimal method that provides similar results to optimal procedures without investing a large computational effort. Also, Simuta (2012) compared the successive inclusion method and a genetic algorithm optimization method in the design of a monitoring network using an objective function similar to the presented in this paper, obtaining similar results with a substantial reduction in the computational effort for the former method.

The criterion used in this research to determine the total number of spatial locations to be included in the OMN is based on the results of the joint total variance and considering the priority order assigned by the method.

To define the number of spatial locations of the monitoring network, we first calculate the square root mean joint total variance (SRMJT).

$$\mathrm{SRMJT} = \sqrt{\frac{\sigma^2_{\mathrm{JT}}}{Nep}} \tag{11}$$

It can be shown that $\sqrt{\frac{\sigma^2_{JT}}{Nep}} \geq \frac{\sigma_{JT}}{Nep}$ for $Nep \geq 1$ and $\sigma^2_{JT} \geq 0$. This means that the SRMJT is an upper bound for the joint average standard error in the spatial estimation grid, which can help us to consider a conservative design scenario.

The criterion consists in selecting the spatial locations where a 90 % of the maximum possible reduction (MPR) of the SRMJT is achieved.

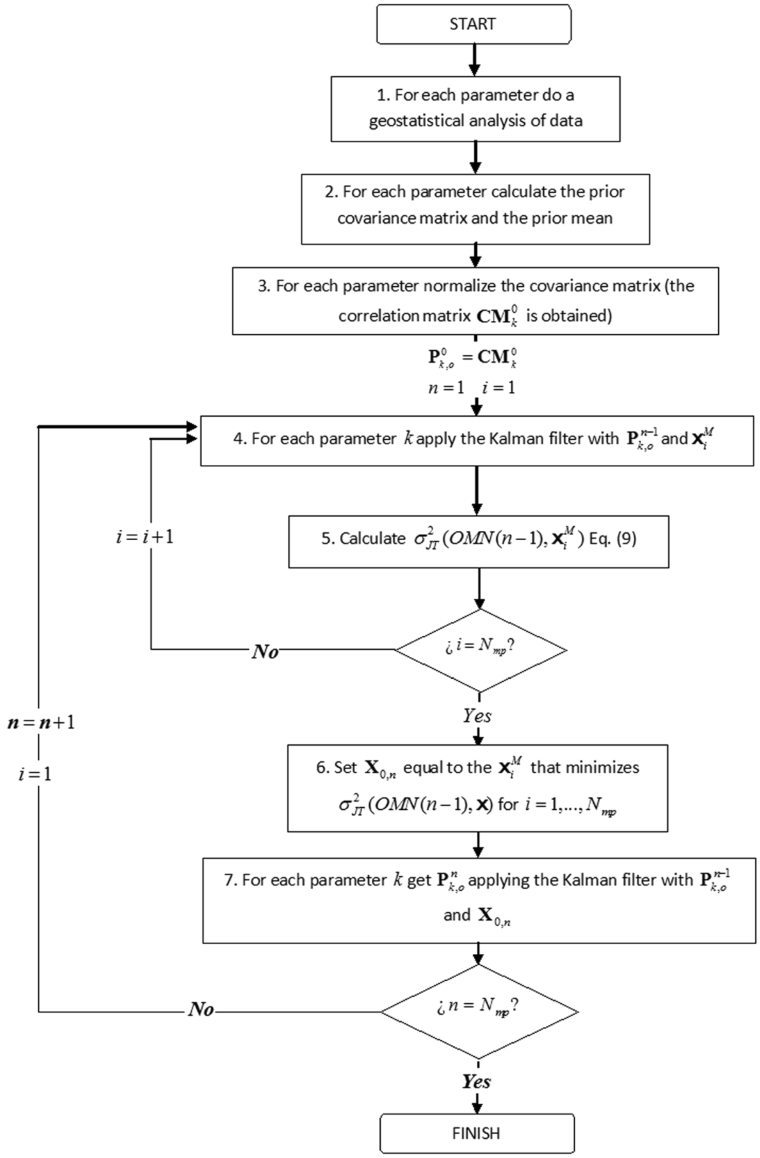

Fig. 1 Flowchart of the proposed methodology

$$\mathrm{MPR} = \mathrm{Max}\sqrt{\frac{\sigma_{JT}^2}{Nep}} - \mathrm{Min}\sqrt{\frac{\sigma_{JT}^2}{Nep}} \tag{12}$$

This way, we guarantee to select a monitoring network that reaches at least the 90 % of the maximum possible reduction for the joint average standard error of the WQIP.

Study area

The proposed methodology is demonstrated by applying it to the design of a low-cost groundwater quality monitoring network for the *Irapuato-Valle* aquifer located in the southern part of the Mexican state of *Guanajuato*. This aquifer represents the main source of water supply for the population of the zone since the *Lerma* River (main surface water source in the study area) suffered a serious decline in the water quality and the quantity of its flow rate as a result of anthropogenic, industrial, and agricultural activities in the region in recent decades.

The *Irapuato-Valle* aquifer has an extent of approximately 2500 km^2, encompassing the cities of *Irapuato*, *Salamanca*, *Valle de Santiago*, *Pueblo Nuevo*, and *Jaral del Progreso*. The deepest zones of the aquifer are about 500 m. It is located between parallels 20° 15′ 00″ and 20° 53′ 00″ N latitude and between meridians 101° 32′ 00″ and 100° 53′ 17″ W longitude (Fig. 2). It is part of the *Lerma* River–*Salamanca* watershed, within the *Lerma-Santiago-Pacífico* administrative region. The highest elevation within the study area is the *Culiacán* Hill, with an altitude of 2830 m above sea level (asl). The climate in the area is warm, with summer rains and an average annual temperature of 18 °C.

The main economic activities in the region are agriculture and industry. Some pollution problems in the region have been produced by inadequate urban wastewater management, abrupt increase in industrial parks, thermal activity, and agricultural activities.

Conceptual hydrodynamic model

For the description of the conceptual hydrodynamic model, information from Water and Sanitation Commission of the State of *Guanajuato* (CEASG) (1998) was used. The uniformity of the superficial geology suggests a relatively homogenous subterranean geological framework. Igneous and sedimentary materials constitute a heterogeneous aquifer unit whose transmissive capacity is mainly associated to the fractured rocks, meanwhile storage is due to the enormous volume of high-porosity materials. The spatial distribution of the lithological units grouped in porous (first two hydrostratigraphic units) and fractured (the third and fourth hydrostratigraphic units) media integrate a heterogeneous unconfined

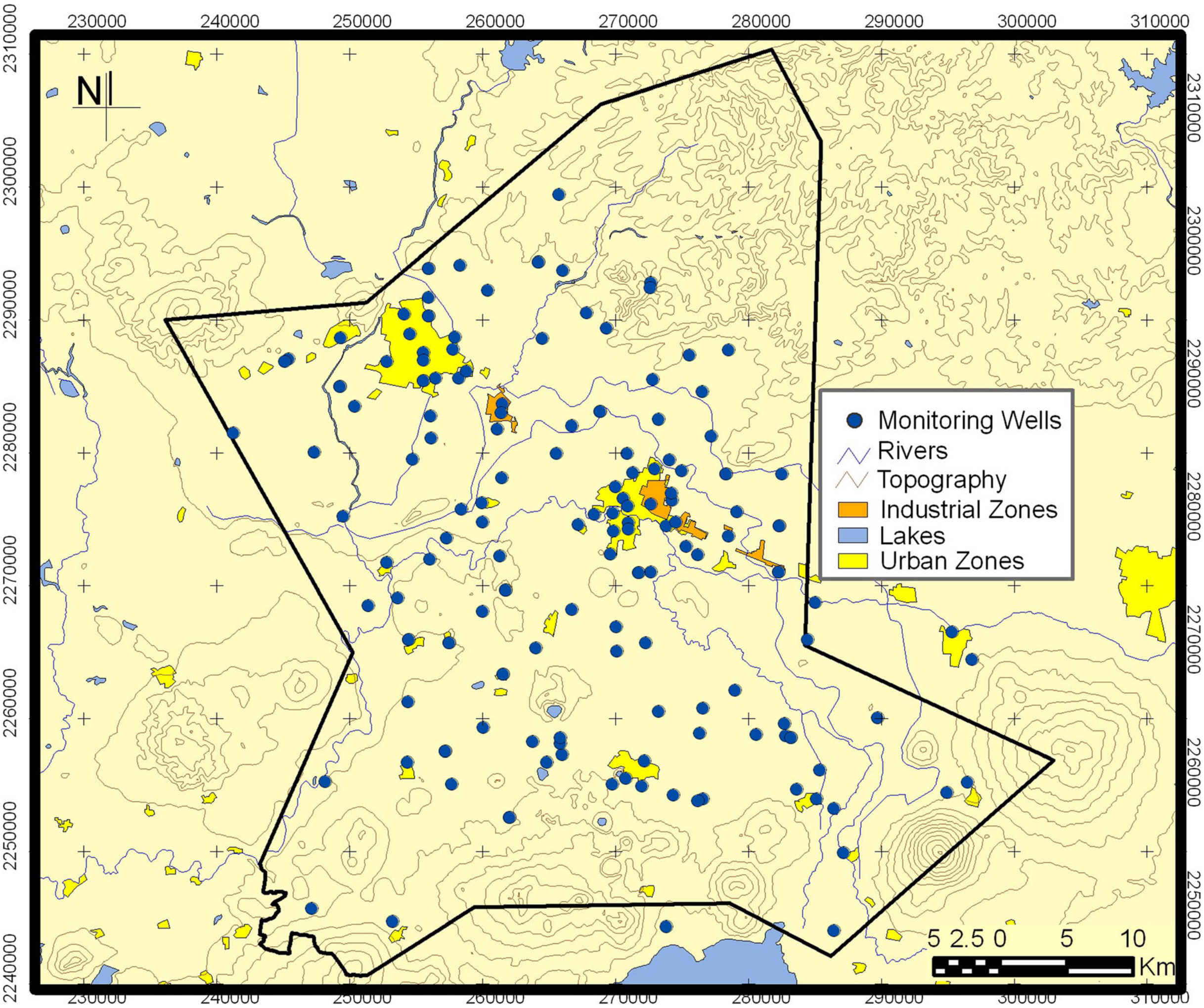

Fig. 2 Geographic location and pilot monitoring network of the *Irapuato-Valle* aquifer

aquifer system that controls the groundwater movement within the study zone.

Three main flow systems have been identified, local, intermediate, and regional, circulating in four hydrostratigraphic units which are described as follows:

> Hydrostratigraphic unit 1, 0 to 150 m thick, consists of intercalated gravel, sands, argillaceous tuffs, and small basaltic lava flows.
> Hydrostratigraphic unit 2, 150 to 250 m thick, composed of sand, gravel, and clayed tuff.
> Hydrostratigraphic unit 3, 150 to 200 m thick, integrated mainly of ignimbrite rocks interbedded with medium to coarse sand.
> Hydrostratigraphic unit 4, greater than or equal to 250 m thick, consisting of altered basalts and medium to coarse sand.

The local flow system occurs in the first hydrostratigraphic unit, the intermediate flow system circulates within the first, second, and third hydrostratigraphic units, meanwhile the regional flow crosses the third and fourth units. The basement is located below 900 m depth, with non-differentiable thickness.

The granular media include mainly alluvial deposits, residual soils, and volcano-sedimentary material, and the fractured media is constituted mainly of basaltic lava flows and interbedded pyroclastic materials.

At initial conditions, the main source of recharge to the *Irapuato-Valle* aquifer occurred from rainfall infiltration on permeable surfaces, through local fractures. Under these conditions, the groundwater flow moved horizontally from east to west in the unconsolidated granular fills and in the fractured medium following the *Lerma* River orientation, being this the principal component of groundwater direction. Below, we describe the three main flow systems identified in the aquifer.

Local flow system. According to the available information, this system is located between 6 and 10 m depth; its upper boundary is constituted by porous or fractured media.

Intermediate flow system. It is recharged mainly from the top of the *Sierra de las Codornices* mountain chain and the *Siete Luminarias* volcanic front. Almost all the water abstraction in the valley takes place on this system. The average groundwater level varies between 25 and 80 m depth.

Regional flow system. Chemical indicators of water from wells allow inferring that the regional flow system circulates in the fractured media of the fourth hydrostratigraphic unit (from top to bottom) as shown in Fig. 3.

Nowadays, groundwater discharge takes place as surface runoff of the *Lerma* River, groundwater flow to the *Pénjamo-Abasolo* Valley, and abstraction from pumping wells. At the southeastern part of the study area, the most significant groundwater inputs to the aquifer occur; some infiltration is added from the *Lerma* River; and at the east side, there is another groundwater input coming from the *Celaya* Valley.

Figure 3 shows a diagram of the conceptual hydrodynamic model of the aquifer in a cross section with northwest-southeast direction, including the defined four hydrostratigraphic units and flow systems, and the cities of *Irapuato* and *Salamanca*.

Groundwater flow directions and hydraulic head evolution

Hydraulic head, topography, and recharge and discharge zones help to interpret the groundwater flow directions. Water moves from higher to lower hydraulic head areas. Topographically, the *Celaya* and *Salvatierra* Valleys are located, respectively, 50 and 100 m higher than the *Irapuato* Valley producing a natural preferential flow direction from the northeast, east, and southeast to the *Irapuato-Valle* aquifer. The main recharge area occurs at the southeastern portion in the direction of the Lerma River. The discharge occurs at the west and southwest of the valley to the *Penjamo-Abasolo* aquifer.

In 1979, the hydraulic head was considered more or less uniform at the center of the valley with a level of 1700 m asl; on the other hand, the southeast portion (current recharge zone) presented a hydraulic head of 1715 m asl. The piezometric analysis for July and December 1998 showed that the hydraulic gradients at the southeast, north, and northeast portions favor the flow to the valley. Two depression cones were identified in the cities of *Salamanca* and *Irapuato* (with hydraulic head elevations of 1640 and 1660 m asl, respectively). Equipotential curves in the rest of the aquifer had an average elevation between 1670 and 1680 m asl. In June 2003, a set of depression cones were identified. The most notable was located from the industrial area to the city of *Irapuato* (hydraulic head of 1635 m asl). On the

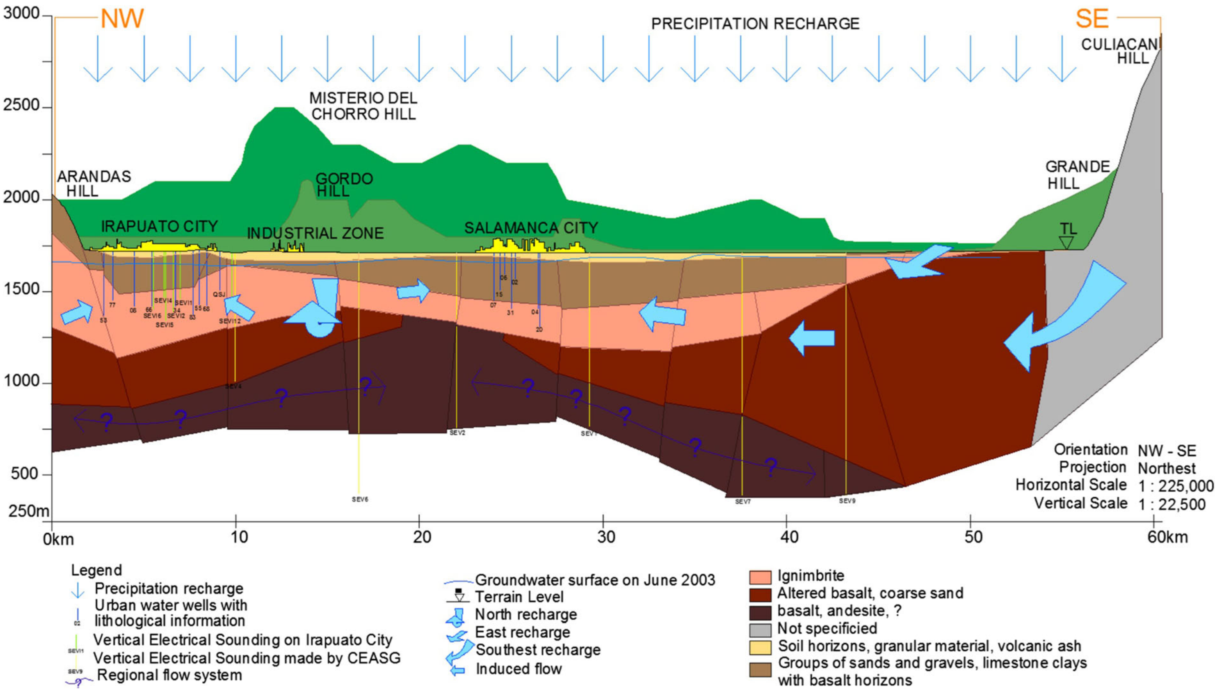

Fig. 3 Hydrodynamic conceptual model of *Irapuato-Valle* aquifer (cross section)

other hand, a dome was produced between the cities of *Irapuato* and *Salamanca* with the highest hydraulic head of 1700 m asl.

At the moment of the sampling campaign carried out in this study, the preferential flow direction was from north to south starting in *San Nicolas Temascatío* with a hydraulic head elevation of 1740 m asl. Several depression cones were formed at the center of the valley. The main entrances of groundwater flow to the valley are at the north with a hydraulic elevation of 1740 m asl, at the northeast and east from the *Celaya* Valley starting with elevations of 1690 m asl, and an additional originated nearby *Valle de Santiago* at the southeast with the highest hydraulic head elevation of 1680 m asl.

Hydrochemistry and groundwater quality

In natural condition previous anthropogenic impact, groundwater flow systems developed from hydraulic gradients along time, producing trends in groundwater chemistry. The typical water chemistry composition in recharge areas is generally associated with low-salinity HCO_3-Ca-type water, following to HCO_3-mix and HCO_3-Na in volcanic terrain. Low-salinity (average 650-mg/l total dissolved solids) HCO_3-Ca (9 % of the total samples) type water was identified close to the highland areas identified as main recharge zones in the north and around Irapuato. However, most of the samples are HCO_3-Na (41 %) and HCO_3-mix (41 %) type, representing the geochemical evolution (average 850-mg/l total dissolved solids) of HCO_3-Ca type interacting with volcanic rocks and basin fill sediments. These main water types representing up to 90 % of total samples have a major composition derived mainly for hydrolysis, dissolution, and cation exchange reactions with volcanic rocks. Additional water types such as Cl-Mg, mix-Ca, mix-Na, and SO_4-mix made up to a total 9 % of the samples; they represent impacted groundwater by wastewater infiltration. HCO_3-Na-type water in the Irapuato-Salamanca region is derived from groundwater interaction with felsic volcanic rocks and/or associated basin fill sediments, whereas HCO_3-mix type in the Valle de Santiago area results from interaction with mafic volcanic rocks. Most of the high arsenic and fluoride waters are associated with the HCO_3-Na type, indicating that leaching from volcanic rocks with rhyolitic composition is a major control.

An analysis of groundwater quality was conducted using data obtained from the pilot well (PW) and historical information of the study area. The most important results are summarized below; details can be seen in González et al. (2003).

Adverse modifications to the environment due to the implementation of intensive crops that may be affecting groundwater quality are related to excessive irrigation and the improper use of agrochemicals (fertilizers and pesticides) or manure. The cities of Irapuato and Salamanca with its large population and industrial concentration generate significant quantities of wastewater from domestic and industrial use. Untreated wastewater is usually discharged to various tributary streams that circulate through the region. Additionally, this contaminated water is used for irrigation of large landfills, a practice that contributes to the spread of pollution in the subsurface through infiltration from irrigation return flows and unlined channels that carry water. In addition to the major population centers, some smaller populations contribute to the generation of liquid waste that sometimes is handled improperly, especially when there is no urban drainage and effluents are discharged through septic tanks.

Results of physical and chemical groundwater analyses were compared with the NOM-127-SSA1-1994. Taking into account major ions like salinity, chloride, sulfate, sodium, and total hardness, groundwater showed no major problems. For example, in 95.5 % of the wells, salinity concentrations were found to be below the permissible limit. For chloride, only the central portion of the aquifer was identified with concentrations exceeding the established limit. For minor elements, fluoride concentration exceeds the permissible limit in the northern part of the aquifer, mainly between *Irapuato* and *Salamanca*. The highest concentrations of fluoride (up to 13 mg/l) were found in deep wells of *Salamanca* that supply water to a thermoelectric power plant. The presence of fluoride in groundwater seems to be associated with natural sources; in addition, high-temperature values facilitate the mobilization of fluoride from rocks to groundwater (high-fluoride values are correlated with groundwater temperature greater than 30 °C).

When analyzing nitrate concentrations in groundwater, two points were detected with values above the permissible limit of 10 mg/l ($N-NO_3$); both are located nearby *Cortazar* and *Valle de Santiago*. Additionally, when oxidizing conditions prevail in the subsurface, such as in the *Irapuato-Valle* aquifer, nitrate is usually a good indicator of the impact of regional wastewater management, but concentrations should be compared with a background value (natural concentration). Nitrate concentrations larger than the background level include the area between the cities of *Irapuato* and *Salamanca*; the south of *Irapuato*; nearby *Pueblo Nuevo*, *Cortazar*, and *Valle de Santiago*; these results represent the impact of agricultural activities on groundwater. These conditions, added to the presence of high levels of total and fecal coliforms in groundwater in some areas of the *Irapuato-Valle* aquifer, should be taken as an alert condition for groundwater quality.

For the analyzed trace elements, the groundwater quality presented some limitations for potable use in certain areas. For example, some arsenic concentrations above the permissible limit were found in the *Salamanca* area and another at the north of *Irapuato*. At least two of these wells belong to the Salamanca Water Supply System (wells 13 CMAPAS and 28 CMAPAS); the others belong to industrial supply wells. Other wells with high levels of arsenic were identified in the *Hoya del Rincón de Parangueo* and *Pueblo Nuevo* region. Considering oxidizing conditions prevailing in most parts of the Irapuato-Valle aquifer, iron values are generally low, and only in three cases, concentrations above the established limit (0.3 mg/l) were identified. Manganese concentrations are low (below 0.05 mg/l) in the section of the aquifer with dissolved oxygen concentrations above 3.5 mg/l; higher concentrations (above 0.05 mg/l) are associated with lower dissolved oxygen concentrations, suggesting a natural redox control. Although the iron and manganese geochemistry are very similar, the latter is somewhat more mobile than the first. Therefore, a larger number of wells (11) exceeded the allowable limit for this parameter (0.15 mg/l). Some of these wells belong to the water supply system of *Salamanca* (2 and 9 CMAPAS), *Valle de Santiago* (7 SAPAM), and *Jaral del Progreso* (2 SMAPAJ). For lead, no high concentration was detected in wells of potable or agricultural use.

Diffuse contamination affecting groundwater quality is basically derived from pollutant sources that may be associated with the following two categories: (i) source discharging substances as a result of lucrative activities (mainly crop irrigation using groundwater and wastewater, pesticide and fertilizer application, and farm waste) and (ii) increase in natural source discharge due to accretion in human activity (surface water-groundwater interaction, natural leaching, etc.).

Historical information was used to understand the groundwater quality evolution due to natural circulation and anthropogenic activities. However, when the methodology is applied, only the most recent sampling campaign is considered because the objective of the geostatistical analysis is to represent the current state of groundwater quality through correlation of the selected indicator parameters. Furthermore, during the optimization of the monitoring network, special emphasis was given to priority zones where an evident degradation of groundwater quality was detected in the last campaign.

Pilot wells

To analyze the aquifer groundwater quality, a set of wells that enables representative sampling of the flow systems identified in the subsurface of the study area was selected (González et al. 2003). It consists of existing active wells that are representative of the different groundwater zones and allow evaluating diffuse and natural pollution through indicator parameters. We call this set of wells, PWs. Most of the wells included in the PW present a screen that captures water from the second and the third hydrostratigraphic units. All of them are representative of the same aquifer system since these units are hydraulically connected, extracting water mainly from the intermediate flow system circulating in the hydrostratigraphic units 1, 2, and 3; nevertheless, four wells capture water from the regional flow system (>400 m depth) circulating in hydrostratigraphic units 3 and 4. PW located in the recharge zones identified in the hydrodynamic model were also selected.

The selection of the PW was done based on constructive characteristics, water quality historical information, ease for measuring and sampling field parameters, potential pollution areas, and zones affected by identified degradation processes. Well depth in the study area range 20–700 m, since wells for drinking water supply are 50 to 300 m depth, with an average of 150 m; the selection of pilot wells considered this along with known nitrate, microorganism, fluoride, arsenic, iron, or manganese concentrations identified from previous reports (unpublished data from Comité Municipal de Agua Potable y Alcantarillado de Salamanca, Valle de Santiago and Irapuato, 2001 and 2003). A field survey to recognize local well conditions, pumping schedule, and irrigation areas with raw wastewater was very useful to identify those with the optimal operational conditions for water sampling.

One field campaign was conducted to gather water samples from the PW, and the information generated from these samples was used for the geostatistical analysis of data necessary in the design of the OMN. Moreover, a subset of the PW was chosen to form the OMN.

The sampling sites were chosen among 145 wells reported in a hydrogeological study conducted by the extinct Ministry of Agriculture (SARH 1979), 135 wells from a study of the Water and Sanitation Commission of the State of *Guanajuato* (CEASG 1998), and wells that contain pollutants that exceed the Official Mexican Standard NOM-127-SSA1-1994 ("Environmental health, water for human use and consumption—and permissible limits of quality and treatments that must be applied for water purification"), reported by the Ministry of Health of the State of *Guanajuato*. The wells of the PW were located geographically with a GPS (NAD27 datum). The PW was formed by 140 wells distributed in the study area. Its spatial distribution can be seen in Fig. 2.

Monitoring network objectives and selected parameters

The objective of the *Irapuato-Valle* aquifer monitoring network is to obtain a general reconnaissance of the water quality, including water types, water origin, and first indications of contamination. To achieve this objective, Jousma (2008) suggests to analyze major ions (Ca, Mg, Na, K, NH_4, Fe, Mn, SiO_2, HCO_3, SO_4, Cl, NO_3, and PO_4) and direct measurements in the field of pH, EC, and temperature. Depending on the hydrogeological framework, Jousma also suggests to sample parameters like As and F. This objective is reflected in NOM-127-SSA1-1994, which was also considered in the selection of the parameters. At each pilot well, water samples were collected from a sampling campaign carried out in December 2003, for laboratory measurement of 20 physicochemical parameters (total dissolved solids, calcium, magnesium, sodium, potassium, bicarbonate, fluoride, sulfate, chloride, nitrate, nitrite, arsenic, total hardness, iron, manganese, lead, phosphate, phenols, total, and fecal coliforms), and in situ measurements were made of six field parameters (conductivity, water temperature, pH, redox potential, dissolved oxygen, and alkalinity).

For the selection of the optimal monitoring sites, it was sought to obtain a low-uncertainty estimate of the selected WQIP for the entire aquifer and with more certainty in priority zones.

Priority zones

Impact of diffuse contamination produced by irrigation with raw wastewater on groundwater has been previously evaluated using a combination of indicators such as chloride, sulfate, nitrate, phosphate, and microorganisms in *San Luis Potosi*, *México* (Cardona et al. 2008). Geogenic sources releasing fluoride, arsenic, and uranium to groundwater have been also identified in the *San Luis Potosi* region (Carrillo-Rivera et al. 2002; Banning et al. 2012). Several similarities can be recognized between *San Luis Potosi* and *Irapuato-Valle* aquifers, tertiary igneous rocks affected by normal faults, basin fill sediments, and diffuse contamination from agricultural practices. The priority zones were delineated for the *Irapuato-Valle* aquifer recognizing areas with concentrations above natural baseline by indicators for anthropogenic contaminant and geogenic sources.

Those regions of lower salinity at the recharge zones present nitrate values lower than 10 mg/l, so this value probably represents the upper limit of concentrations that can arise naturally in the Irapuato-Valle aquifer, and therefore, this value can be considered the level of geochemical background for nitrate. The regions where nitrate concentrations are higher than the background level like the area between the cities of Irapuato and Salamanca; the area at the south of Irapuato City; and near the towns of Pueblo Nuevo, Valle de Santiago, and Cortazar represent the impact of agricultural activities on groundwater. In addition, nitrate concentrations below 250 m have an upper value of about 10 mg/l, representing groundwater with low anthropogenic impact (Fig. 4).

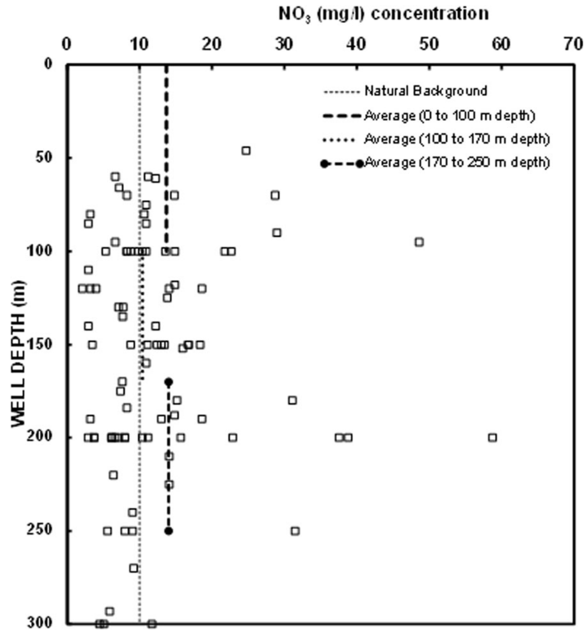

Fig. 4 Nitrate concentration versus well depth

Results and discussions

Once the monitoring objectives and the groundwater quality conceptual model were defined, and the sampling campaign of the most representative groundwater quality parameters in the PW was conducted, the design of the optimal monitoring network was started.

Geostatistical analysis of selected parameters

In the geostatistical analysis for the monitoring network design, we used the data collected in 2003 from the 140 PW. The geostatistical analysis was done for arsenic, chloride, electrical conductivity, fluoride, manganese nitrate, sodium, temperature, and hardness, because only these parameters had enough data above the detection limit. It was necessary to apply a natural logarithmic transformation to all parameter data in order to accomplish the normal distribution hypothesis required in a geostatistical analysis. The selected variogram models for each WQIP are shown in Table 1.

A covariance matrix was then calculated for each parameter except for hardness for which a power model (unbounded) was adjusted. Therefore, this parameter was not included in the OMN design.

Estimation grid

As mentioned before, the statistical objective of the design was to select the locations needed to obtain a low-uncertainty estimate of the WQIP for the whole aquifer and with lesser uncertainty in priority zones for the most representative parameters.

To achieve this, we first defined a preliminary spatial estimation grid, with squared elements of 2 km side length, and priority zones for each WQIP in auxiliary grids composed by squared elements of 1 km side length to assign them a higher weight in the optimization process; i.e., the monitoring points selected inside or nearby priority zones have a higher influence over these auxiliary grids than monitoring points selected away from them. This fact provide them additional advantage

Table 1 Variogram models of water quality parameters

Variable	Model	Nugget	Sill	Range	Akaike criterion
Ln arsenic	Spherical	0.3	0.9	25,000	−660.373
Ln chloride	Spherical	0.2	0.75	17,000	−848.031
Ln conductivity	Spherical	0.05	0.16	30,000	−1289.045
Ln fluoride	Spherical	0.19	0.56	13,000	−894.810
Ln manganese	Spherical	0.24	0.38	20,000	−567.263
Ln nitrate	Spherical	0.35	0.40	26,000	−838.986
Ln sodium	Spherical	0.14	0.64	15,500	−939.973
Ln temperature	Spherical	0.003	0.0235	9,000	−1630.710
Ln hardness	Power	0.0	0.258	0.043	−620.088

in reducing the joint total variance since more estimation points are affected during the optimization; therefore, they have a higher probability to be chosen. As an example, the preliminary estimation grid for arsenic in blue circles and the corresponding auxiliary grid in red triangles are shown in Fig. 5.

The union of the positions of all the auxiliary grids and the positions of all the preliminary spatial estimation grids constitute the estimation grid with 1698 nodes (Fig. 6); the 140 positions of wells of the PW are the possible monitoring locations.

Applying Eq. 6, the prior spatial covariance matrix for each WQIP was calculated from its variogram model (Table 1). The covariance matrix for each parameter was constructed for 1698 (estimation points) + 140 (monitoring points) = 1838 positions, so its dimension was of 1838 columns by 1838 rows. The elements of the prior covariance matrix of each WQIP that correspond to the auxiliary grid are non-zero only if they are within its priority zones. In this way, the priority zones are considered in the optimization procedure only for the corresponding parameter. Afterward, the correlation matrix for each parameter was calculated as was explained in Materials and methods section.

Priority order

When the optimization method explained before is applied, the order in which the spatial locations are selected represents a priority order because the location that reduces the most variance (in other words, the location that gives maximum information) is selected at each round. In this way, the selection order is an indicator of how important are the data obtained at those locations in reducing the joint total variance; the smaller the selection order, the bigger the priority. It can be seen clearly in Fig. 7a that the joint total variance is greatly reduced when choosing the first spatial locations, but as the number of selected locations increases, the amount of joint total variance reduction decreases.

Criterion for determining the total number of monitoring locations

Following the chosen criterion to determine the total number of monitoring locations in the monitoring network, it can be seen in Fig. 7b that after monitoring the 69 locations with larger priority, 90 % of the MPR is achieved.

If we apply the optimization procedure for each WQIP separately, with the 69 monitoring wells with higher priority, we obtain a 93.90 % of the MPR for arsenic, 92.22 for chloride, 95.03 for electrical conductivity, 87.92 for fluoride, 87.29 for manganese, 77.86 for nitrate, 92.02 for sodium, and 88.15 for temperature. For each WQIP, 90 % of its MPR will be achieved at 55 sampling locations for arsenic, 62 for chloride, 48 for electrical conductivity, 75 for fluoride, 77 for manganese, 97 for nitrate, 63 for sodium, and 74 for temperature.

Even though when 69 positions are monitored, the MPR values for nitrate, manganese, fluoride, and temperature are lower than 90 %; these values are considered acceptable. When this value is not accepted for a certain parameter, additional monitoring positions should be included in the monitoring network following the priority order criteria.

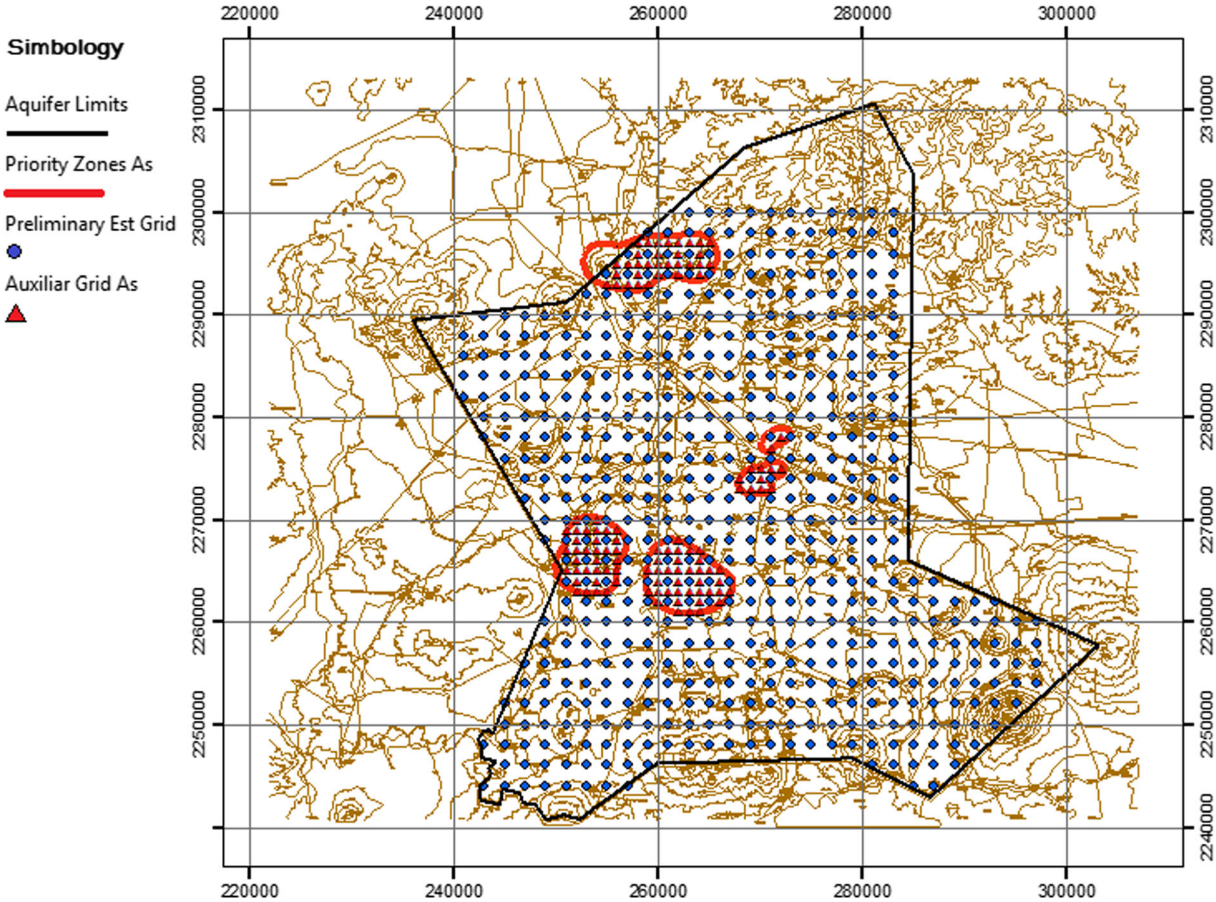

Fig. 5 Preliminary estimation grid (*blue circles*) and auxiliary estimation grid for arsenic (*red triangles*)

Optimal monitoring network

In this problem and according to what was exposed before, 69 monitoring points were selected to constitute the OMN. Figure 8 shows the priority order for the selected wells to constitute the OMN. The minimum depth for the monitoring wells is 60 m and the maximum is 700 m; the average depth is 168 m with a standard deviation of 108 m. The screen at these wells usually starts about 20 m below the ground surface, and it is maintained through all the well depths. Thirty wells capture water from the hydrostratigraphic unit 1 (≤150 m) while 20 wells from the two upper hydrostratigraphic units (≤400 m). One well reaches the hydrostratigraphic unit 3 (525 m depth) and one the hydrostratigraphic unit 4 (700 m depth). The two wells with the largest known depth capture water from both the intermediate and regional flow systems; the others capture water from the intermediate flow system only. For 17 wells, the depth is unknown; however, according to the hydrogeochemical characteristics of the abstracted water, it is provided from the intermediate flow system in 16 of these wells; the other wells additionally capture water from the regional flow system.

Water quality indicator parameter estimation

As mentioned before, the Kalman filter needs prior estimates of the state vector and the estimate error covariance matrix. For each parameter, the state vector was estimated with the average of natural logarithm transformation of data corresponding to positions with water quality data measured in the field. After applying the Kalman filter, estimates were transformed to the original variable with the formulas presented in Journel and Huijbregts (1978) for lognormal kriging.

As an illustration of the estimates that would be obtained using the optimal monitoring network, WQIP estimates were obtained on the estimation grid with the available data from December of 2003.

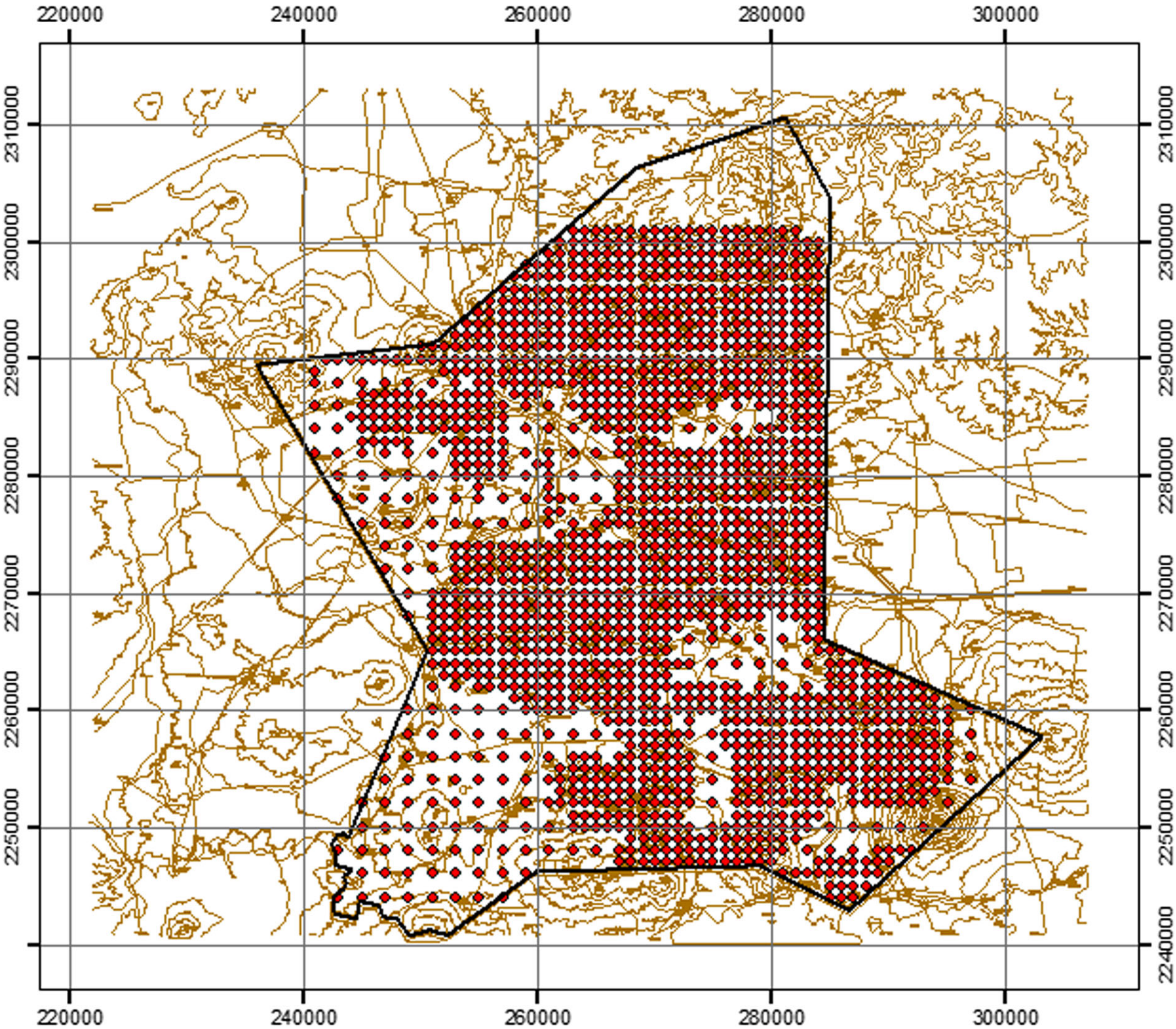

Fig. 6 Estimation grid

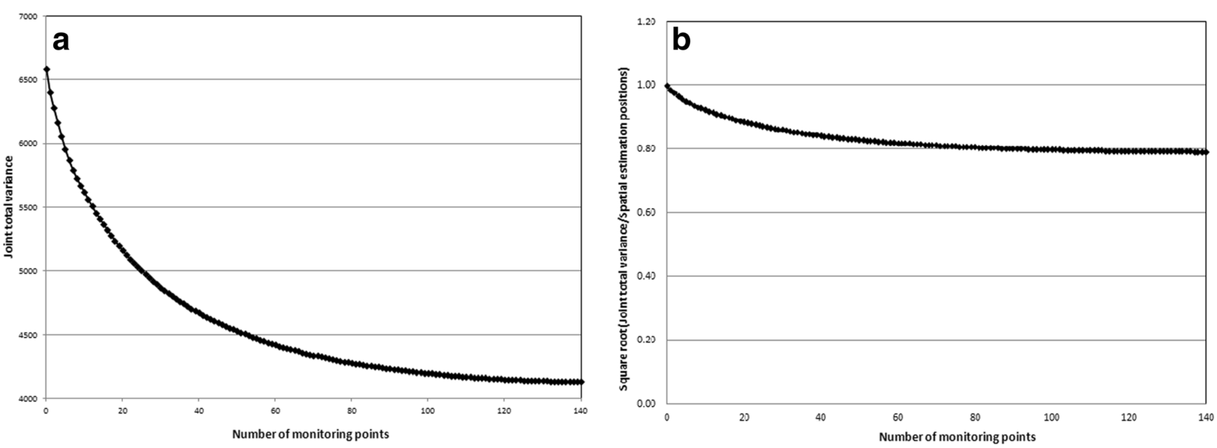

Fig. 7 **a** Joint total variance versus number of spatial monitoring points and **b** square root (joint total variance/spatial estimation positions) versus number of spatial monitoring points

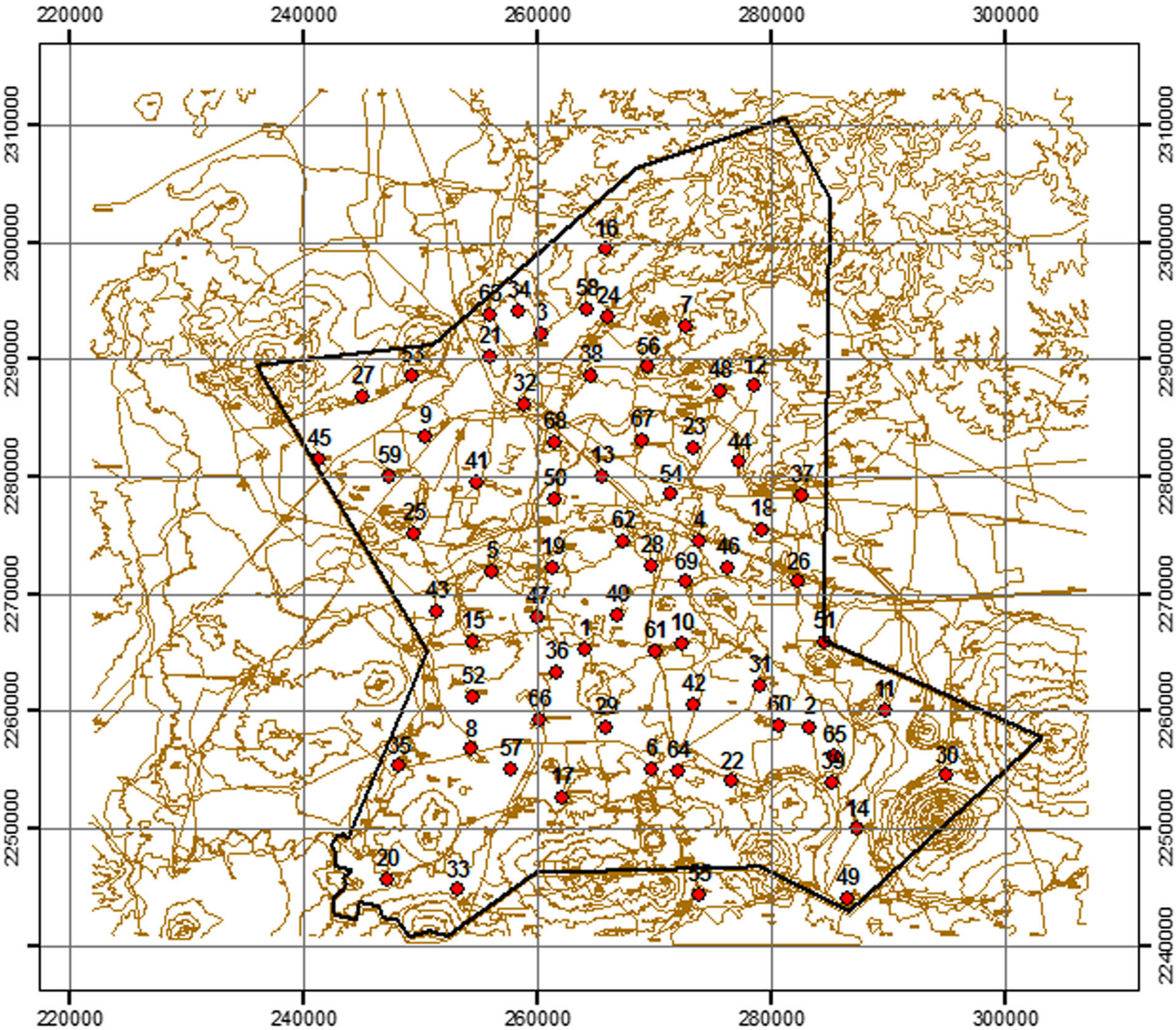

Fig. 8 Priority order for the wells of the OMN

Figures 9 and 10 show arsenic and temperature estimates obtained with the proposed monitoring network. Figures 11 and 12 show their respectives estimate error variances; it can be seen that low values are obtained inside the priority zones. The variance maps together with practical criteria can be very useful in choosing new monitoring locations in the zones with the maximum variances.

To compare the estimates obtained with the OMN (69 wells) and the PW (140 wells), some statistics of differences between estimates and error variances obtained with the KF for both cases are presented in Table 2. To have an idea of the estimate error magnitudes, database mean values of measured values are also shown. In the column of PW-OMN estimates, the mean absolute differences are shown. It can be seen that differences between estimates for both options are small compared to the database mean values (range between 0.11 and 5.17 %), except for fluoride (64.29 %) and chloride (13.88 %). The large relative differences in fluoride estimates for both monitoring options occur probably because the natural logarithm-transformed data of this parameter (employed to construct the variogram model) present the largest coefficient of variation, and coincidently, some outliers (the two largest values for this parameter) are not included in the OMN; on the other hand, differences in chloride estimates are

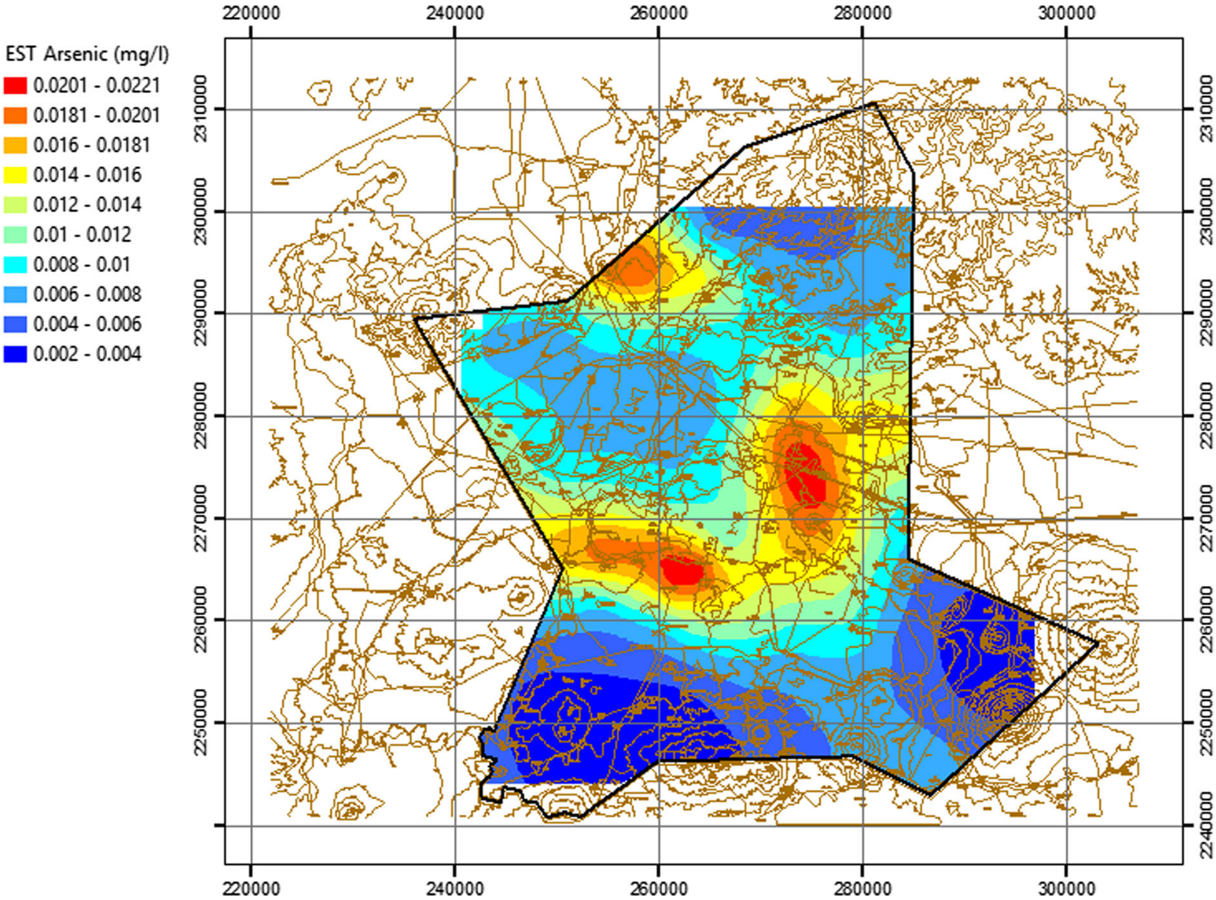

Fig. 9 Arsenic estimate (mg/l)

due to the contribution of an outlier (largest value for this parameter) selected by the OMN that is surrounded by wells with considerable lower values that are included only in the PW. The square root mean variance (SRMV) is a measure of the average uncertainty in the estimation grid when employing the OMN or the PW. The comparison in Table 2 shows the increase in the average uncertainty when employing the OMN and the PW; it can be seen that when using the OMN, a marginal loose of certainty is obtained compared to the PW.

Excluding travel times, the average cost in 2014 for sampling and laboratory determination of the analyzed parameters for each well is 267.25 US dollars; therefore, the total cost for monitoring the PW would be 37,415.00 dollars, and for the OMN, would be 18,440.95. In this sense, adopting the proposed OMN would represent savings for at least 18,974.05 dollars, by accepting the loss of information indicated in Table 2.

Discussion

Initial developments of the methodology introduced in this paper were presented in Herrera et al. 2004 and Júnez 2005. In those research, the monitoring network design included two steps. In the first step, a preliminary sampling network was obtained without considering the priority zones. To complement this sampling network, additional wells were selected by inspection in order to get a maximum increase of 10 % of the estimate standard error (ESE) within the priority zones, compared to the ESE using all the pilot wells. The automated procedure presented in this paper has the advantage of including the priority zones during the optimization in one single step, eliminating the need to do the second step by hand.

Another important difference between the methodology presented in Herrera et al. 2004 and Júnez 2005 and the methodology presented in this paper is that in the

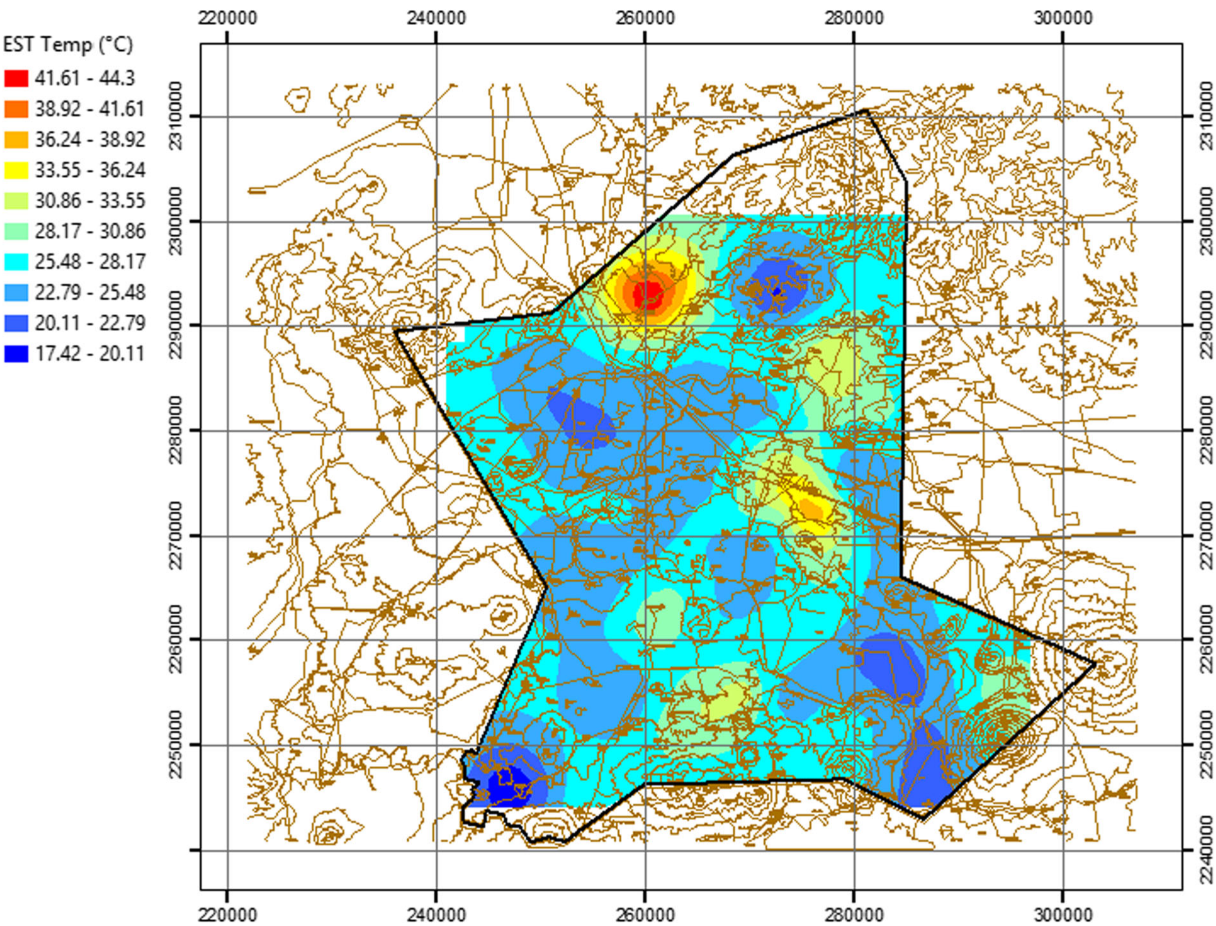

Fig. 10 Temperature estimate (°C)

former, the size of the sampling network was defined by a visual approach, meanwhile in the latter, a quantitative criterion that allows to control the level of expected estimate uncertainty was used.

Hergt (2009) developed two different monitoring designs, both applying initial developments of the methodology introduced in this paper (Herrera et al. 2004; Júnez 2005), calculating for one case the covariance matrices from a geostatistical analysis of the most significant water quality parameters resulting from a factor analysis and for the other, using the covariance matrix of the factor analysis itself. Similar results were obtained in both cases.

If the optimization procedure presented in this paper was applied for each WQIP separately, different sets of monitoring positions would be necessary to satisfy the same level of uncertainty for all the parameters since spatial correlation is usually different for each WQIP; therefore, several redundant monitoring positions could be obtained for a global monitoring network that enclosed the different sets; i.e., this type of design could represent large and/or unnecessary groundwater quality monitoring costs.

Conclusions and recommendations

The proposed methodology considers various water quality parameters and priority zones (where certain parameter values were exceeded for each parameter) simultaneously in the design of an optimal monitoring network. It employs the spatial correlation between data of each parameter using spatial covariance models derived from geostatistical analysis.

An underlying hypothesis of the proposed methodology is that the optimal monitoring network will be useful if aquifer conditions do not change dramatically (for example, in cases where there are no significant changes in land use or in groundwater extraction). We

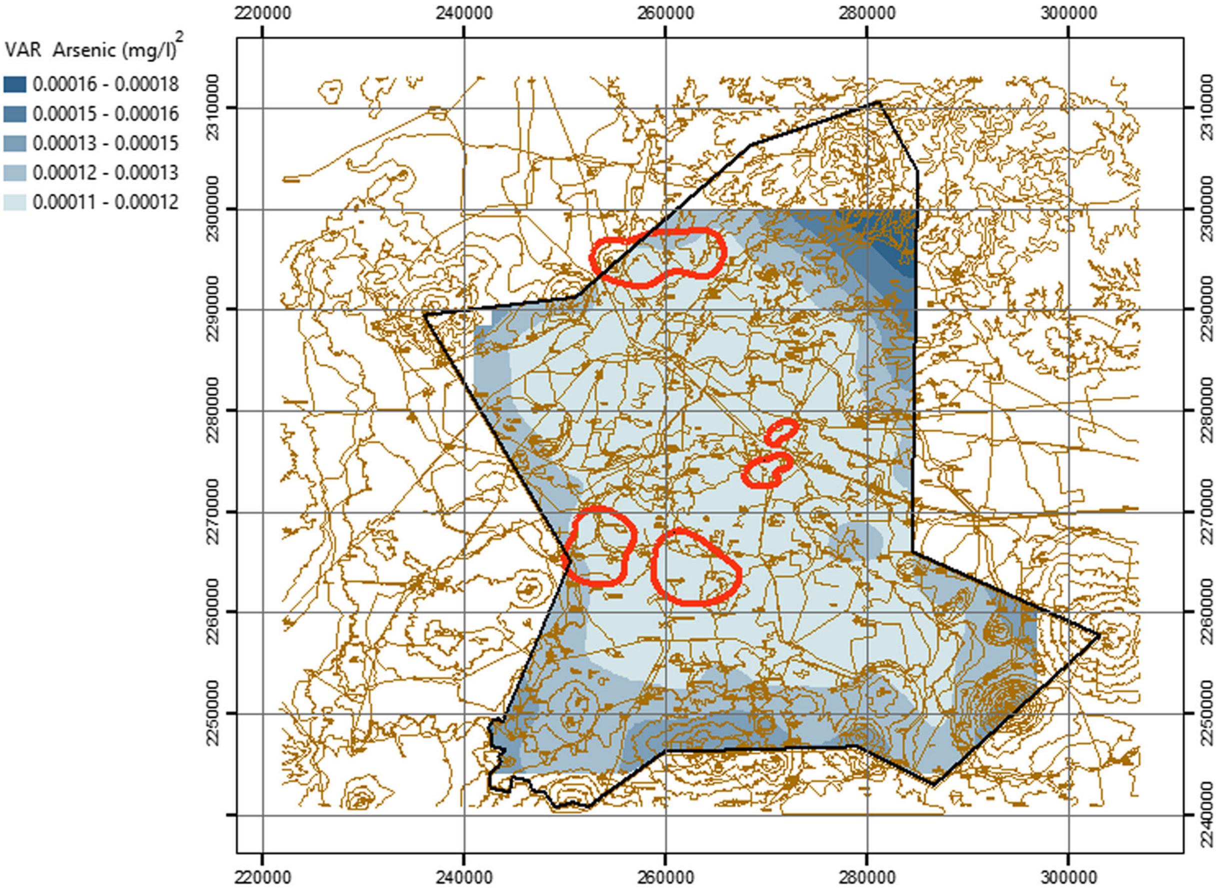

Fig. 11 Estimate error variances and priority zones for arsenic

recommend performing a large sampling campaign and redesigning the proposed monitoring network periodically. This period will depend on groundwater flow velocities and changes in aquifer conditions.

Since changes in contaminant concentrations depend partially on flow directions, further work should include them to propose the sampling network.

For all the analyzed water quality parameters, it was necessary to apply a natural logarithmic transformation to data in order to accomplish the normal distribution hypothesis required in a geostatistical analysis.

The optimal monitoring network resulted in 69 of the 140 wells of the set of pilot wells. This implies a 50 % of cost reduction with a marginal increase in uncertainty. Furthermore, a maximum increase of 4.7 % of the SRMV values within all the nodes of the estimation grid was obtained for the selected OMN of 69 wells, compared to the SRMV calculated with the PW of 140 wells.

The estimate error variances were minimized inside the priority zones for all the parameters during the optimization procedure. Positions in or nearby the priority zones were automatically the first to be included in the optimal monitoring network. An important advantage of the proposed methodology is that a large set of wells is selected within the priority zones; therefore, the estimate error variances are smaller at those zones than in the rest of the aquifer. A groundwater quality monitoring network design that considers various WQIPs simultaneously could reduce groundwater quality characterization costs substantially.

Areas with large values of total variance after measuring in the monitoring locations indicate that the spatial distribution of that parameter is not represented with certainty in those areas. This means that new wells should be added to the optimal monitoring network in those zones. To propose the locations of the new wells, potential pollution sources and hydrogeological criteria should be also considered.

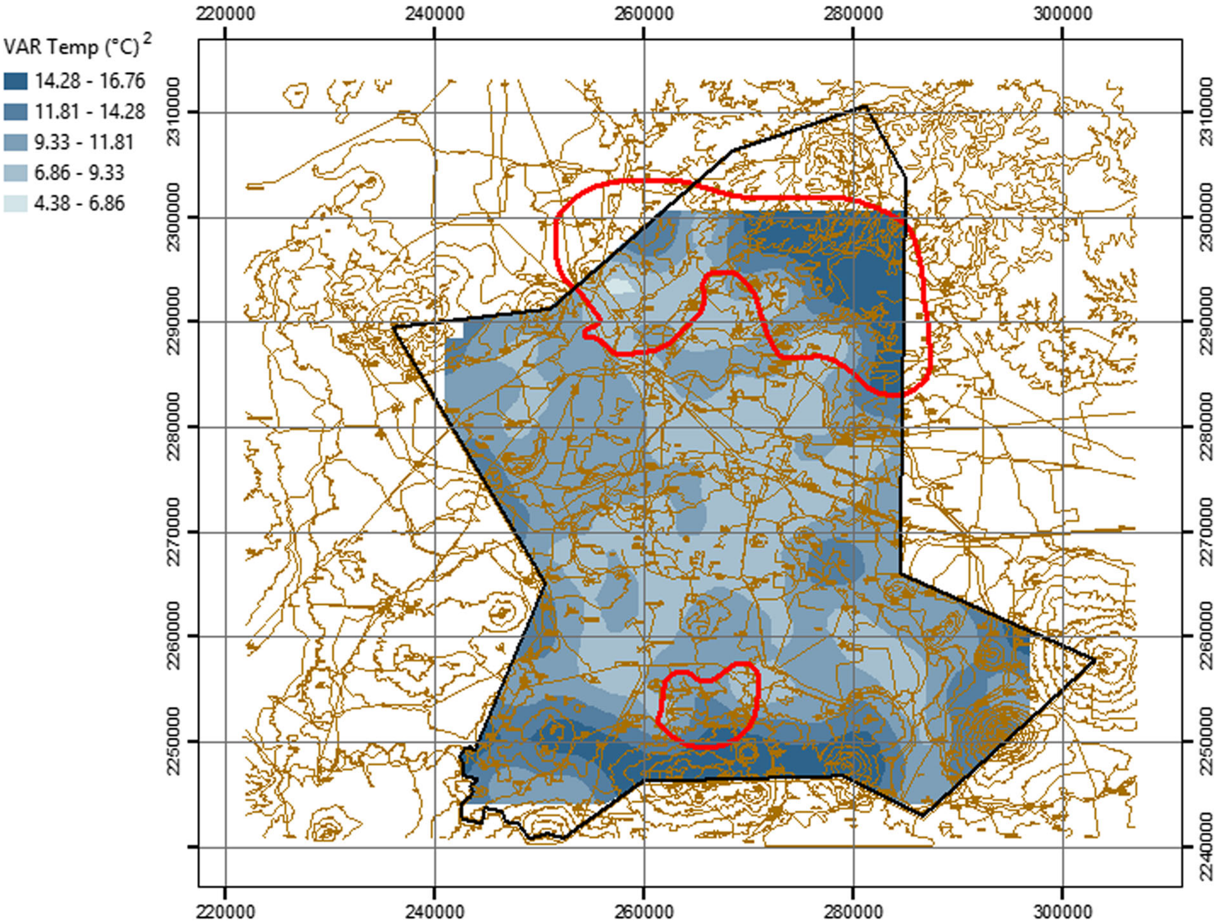

Fig. 12 Estimate error variances and priority zones for temperature

An important advantage of the proposed method is that it can be used to define a water quality preliminary monitoring network since it requires only data of the analyzed parameters for a certain sampling campaign unlike entropy theory-based design methods presented in literature that require time series. Furthermore, when

Table 2 Comparison of some statistics of the estimates obtained with the Kalman filter for the data from the pilot wells and the optimal monitoring network

Parameter (units)	Database mean values	Coefficient of variation of data with Ln transformation	PW-OMN estimates (mean absolute differences)	Percent of mean absolute differences with respect to database mean values	SRMV	
					PW	OMN
Arsenic (mg/l)	0.011	0.18	0.00007	0.64	0.011	0.011
Chloride (mg/l)	35.44	0.25	4.92	13.88	29.697	30.513
Electrical conductivity (μS/cm)	848	0.06	43.88	5.17	249.977	256.138
Fluoride (mg/l)	0.98	1.77	0.63	64.29	0.554	0.571
Manganese (mg/l)	0.042	0.32	0.002	4.76	0.004	0.004
Nitrate (mg/l)	11.51	0.78	0.29	2.52	7.655	7.695
Sodium (mg/l)	74.96	0.19	1.56	2.08	57.987	59.751
Temperature (°C)	26.7	0.06	0.03	0.11	3.143	3.29

times series data are available, this method can be extended for a space-time design as the developed by Júnez-Ferreira and Herrera (2013) for hydraulic head monitoring network. Moreover, spatial and/or temporal estimates of various water quality parameters and its uncertainty could be obtained for the designed monitoring network.

The static Kalman filter introduces a smoothing effect in the estimates as kriging methods do. To avoid this problem, we can introduce a correction in the estimation method to be applied once the data were collected from the proposed optimal sampling network. Among other proposals, Yamamoto (2005) introduces a four-step method that uses interpolation variances for correcting the smoothing effect of ordinary kriging estimates. An alternative may be developing an analogue modification of the Kalman filter equations to obtain estimates free of the smoothing effect.

It is well known that boreholes are the proper sites for sampling groundwater; unfortunately, in countries like Mexico, it is uncommon to drill boreholes for monitoring groundwater heads and quality, except for specific studies, due to the high costs that they can represent. Indeed, monitoring campaigns conducted to date in the Irapuato-Valle aquifer are largely based on the interpretation of groundwater samples collected from wells used for irrigation and drinking purposes. These wells extract water from variable depths (average close to 175 m) under different flow rates and operation times; in this manner, the abstracted water represents a mixture of a large extent of the saturated thickness. In order to define the vertical distribution of groundwater quality indicator parameters in the aquifer, it is necessary the drilling and instrumentation of boreholes with specific characteristics to accomplish this purpose. Future works could include uncertainty associated to this kind of samples, specifically in sampling errors.

References

ASCE (American Society of Civil Engineers). (2003). *Long-term groundwater monitoring: the state of the art*. USA: Task Committee on the State of the Art in Long-term Groundwater Monitoring Design.

Banning, A., Cardona, A., & Rüde, T. (2012). Uranium and arsenic dynamics in volcano-sedimentary basins—an exemplary study in north-central Mexico. *Applied Geochemistry, 27*, 2160–2172.

Cardona, A., Carrillo-Rivera J. J., Castro-Larragoitia, G. J., & Graniel-Castro, E. (2008). *Combined use of indicators to evaluate waste water contamination to local flow systems in semi-arid regions: San Luis Potosi, Mexico. Selected Papers Series of the International Association of Hydrogeologists (SPS-IAH) on the theme:* "*Groundwater flow understanding from local to regional scales*". Balkema, Taylor & Francis, 85–104.

Carrillo-Rivera, J. J., Cardona, A., & Edmunds, W. M. (2002). Use of abstraction regime and knowledge of hydrogeological conditions to control high-fluoride concentration in abstracted groundwater: San Luis Potosí basin, México. *Journal of Hydrology, 261*, 24–47.

Chadalavada, S., Datta, B., & Naidu, R. (2011). Uncertainty based optimal monitoring network design for a chlorinated hydrocarbon contaminated site. *Environmental Monitoring and Assessment, 173*, 929–940.

CEASG (Geofísica de Exploraciones GUYSA, S.A. de C.V.). (1998). *Estudio hidrogeológico y modelo matemático del acuífero del Valle de Irapuato—Valle de Santiago*. Documento del plan estatal hidráulico de Guanajuato, México.

Dutta, D., Das Gupta, A., & Ramnarong, V. (1998). Design and optimization of a ground water monitoring system using GIS and multicriteria decision analysis. *Ground Water Monitoring and Remediation, 18*(1), 139–147.

González, L., Herrera, G., Cardona, A., Mora, J., Júnez, H. E., Becerra, L., & Gutiérrez, C. (2003). *Contaminación difusa en el agua subterránea en el acuífero Irapuato-Valle, Gto*. México: Comisión Nacional del Agua, IMTA (Instituto Mexicano de Tecnología del Agua).

Hergt, T. (2009). *Diseño optimizado de redes de monitoreo de la calidad del agua de los sistemas de flujo subterráneo en el acuífero 2411 "San Luis Potosí": Hacia un manejo sustentable*, Tesis de doctorado, Universidad Autónoma de San Luis Potosí.

Herrera, G. S. (1998). *Cost effective groundwater quality sampling network design*, Ph. D. Dissertation, University of Vermont.

Herrera, G. S., & Pinder, G. F. (2005). Space-time optimization of groundwater quality sampling networks. *Water Resources Research, 41*, W12407. 15 pp.

Herrera, G. S., Júnez-Ferreira, H. E., González, L., & Cardona, A. (2004). Diseño de una red de monitoreo de la calidad del agua para el acuífero Irapuato-Valle, Guanajuato. *Memorias del XVIII Congreso Nacional de Hidráulica, AMH, SLP, México*.

Jazwinski, A. H. (1970). *Stochastic processes and filtering theory*. London: Academic.

Journel, A. G., & Huijbregts, C. J. (1978). *Mining geostatistics* (p. 570). London: Academic.

Jousma, G. (2008). *Guideline on groundwater monitoring for general reference purposes, Report GP 2008–1*. The Netherlands: International Groundwater Resources Assessment Center (IGRAC).

Júnez, H. E. (2005). *Diseño de una red de monitoreo de la calidad del agua para el acuífero Irapuato-Valle, Guanajuato*. Tesis de Maestría, UNAM, México. Online access http://132.248.9.195/ptd2012/anteriores/0339152/Index.html.

Júnez-Ferreira, H. E., & Herrera, G. S. (2013). A geostatistical methodology for the optimal design of space-time hydraulic-head monitoring-networks and its application to the Valle de Querétaro aquifer. *Environmental Monitoring and Assessment, 185*(4), 3527–3549.

Kumar, S., Sondhi, S. K., & Phogat, V. (2005). Network design for groundwater level monitoring in Upper Bari Doab canal tract, Punjab, India. *Irrigation and Drainage, 54*, 431–442.

Li, J., Bárdossy, A., Guenni, L., & Liu, M. (2011). A Copula based observation network design approach. *Environmental Modelling & Software, 26*, 1349–1357.

Lin, Y., & Rouhani, S. (2001). Multiple-point variance analysis for optimal adjustment of a monitoring network. *Environmental Monitoring and Assessment, 69*, 239–266.

Masoumi, F., & Kerachian, R. (2010). Optimal redesign of groundwater quality monitoring networks: a case study. *Environmental Monitoring and Assessment, 161*, 247–257.

Preziosi, E., Petrangeli, A. B., & Giuliano, G. (2012). Tailoring groundwater quality monitoring to vulnerability: a GIS procedure for network design. *Environmental Monitoring and Assessment*. doi:10.1007/s10661-012-2826-3.

Samper, F. J., & Carrera, J. (1990). *Geoestadística, aplicaciones a la hidrogeología subterránea*. Barcelona: Centro Internacional de Métodos Numéricos en Ingeniería, Universidad Politécnica de Cataluña.

SARH (Secretaría de Agricultura y Recursos Hidráulicos). (1979). *Estudio geohidrológico cuantitativo de los acuíferos del Alto Lerma, Guanajuato*. Geohidrología Mexicana, S.A., México.

Simuta, R. (2012). *Diseño óptimo de redes de monitoreo de la calidad del agua subterránea con muestreo a diferentes profundidades*. Tesis de Doctorado, UNAM, México (132.248.9.195/ptd2013/junio/500017791).

Yamamoto, J. (2005). Correcting the smoothing effect of ordinary kriging estimates. *Mathematical Geology, 37*(1), 69–94.

Yeh, M. S., Lin, Y. P., & Chang, L. C. H. (2006). Designing an optimal multivariate geostatistical groundwater quality monitoring network using factorial kriging and genetic algorithms. *Environmental Geology, 50*, 101–121.

Zaidi, F. K., Ahmed, S. H., Dewandel, B., & Maréchal, J. (2007). Optimizing a piezometric network in the estimation of the groundwater budget: a case study from a crystalline-rock watershed in southern India. *Hidrogeology Journal, 15*(6), 1131–1145.