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


**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

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

## Introduction

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

Published online: 17 December 2015

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

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

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 $j$th 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

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

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 $\bullet$, 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:

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

where $\sigma_{k,j}^2$ 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 $j$th 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