Linked Optimization Model for Groundwater Monitoring Network Design

Deepesh Singh and Bithin Datta

Abstract Groundwater is a major source of water supply for irrigation, industrial use, and other public consumptions due to its quality, local accessibility, and relative cost. Groundwater resource is also equally vulnerable to contamination and depletion as the surface water. The detection of these contaminants is difficult, as they are not visible as the surface water systems. The detection and monitoring of the contaminants is very much important for the prediction of the contaminant transport process, and for designing efficient remedial measures. This study involves the development of methodologies for optimal design of groundwater contamination monitoring networks for detection the contaminant movement in groundwater systems, based on the application of simulated annealing as an optimization tool, and geostatistical kriging. This methodology is developed for single and time-varying optimal network design for different management periods. A budgetary constraint is used to limit the number of monitoring wells to be installed in a particular management period. The kriging linked simulated annealing (SA)-based optimization model essentially utilizes a numerical flow and transport simulation model (MODFLOW and MT3DMS) to simulate the physical and geochemical processes. It searches for an optimal set of permissible number of monitoring wells. The specified objective function of minimizing the contaminant mass estimation error is found to be quite suitable. The methodologies developed are applied to an illustrative study area comprising of homogeneous unconfined aquifer. The performance of the methodology is evaluated for the illustrative study area and the limited evaluation results demonstrate the potential applicability of the methodologies developed.

D. Singh (✉)
Department of Civil Engineering, Harcourt Butler Technological Institute, Kanpur, U.P., India
e-mail: dr.deepeshsingh@gmail.com

B. Datta
Discipline of Civil Engineering, College of Science Technology and Engineering, James Cook University, Townsville 4811, Australia
e-mail: bithin.datta@jcu.edu.au

A.K. Sarma et al. (eds.), *Urban Hydrology, Watershed Management and Socio-Economic Aspects*, Water Science and Technology Library 73,
DOI 10.1007/978-3-319-40195-9_9

Keywords Groundwater monitoring network design · Simulated annealing · Kriging · Groundwater contamination · Contaminant transport simulation

1 Introduction

Groundwater, under most conditions, is safer and more reliable for use than surface water. Groundwater my also be contaminated by some of the common activities which cause groundwater contamination like: illegal and unmonitored injection of pollutants in the aquifer, leakage from underground tanks, and pipelines carrying sewage and other toxic contaminants; insufficient knowledge about the application of pesticides and fertilizers in agricultural fields, land disposal of wastes, etc. The detection of groundwater contamination is difficult as it is not openly visible unlike surface water systems. Prediction of the transport of contaminants in groundwater systems also becomes difficult due to various associated uncertainties. The detection and remediation of groundwater contamination involves monitoring of the contaminant plume in groundwater systems. This study is aimed at developing methodologies for optimal monitoring of groundwater contamination. A groundwater model can have two distinct components: (i) groundwater flow and (ii) groundwater contaminant transport components. The uncertainties involved in the prediction of plume movement and the economic constraint to limit the number of monitoring well installations necessitates the design of an optimal monitoring network design. The economic efficiency is incorporated by limiting the maximum permissible number of monitoring locations as an upper limit.

Especially during last two decades, the problem of designing an optimal monitoring network under condition of uncertainty has been addressed by number of researchers. Meyer and Brill (1988) developed a method that involves the use of two independent but linked models, a groundwater contaminant transport simulation model and an optimization model. Loaiciga (1989), Loaiciga et al. (1994, 1995) formulated groundwater monitoring (GWM) optimal sampling plan as mixed integer programming (MIP) problem. Meyer and Brill (1988) also utilized MIP to determine optimal location of a network of groundwater monitoring wells under conditions of uncertainty. Datta and Dhiman (1996) developed a groundwater quality monitoring network using a groundwater contamination transport simulation and optimization linked model. Lin et al. (2001) estimated the spatial maps of transmissivity and simulated using ordinary kriging (OK) and simulated annealing. Reed and Minsker (2004) demonstrated the use of high-order Pareto optimization on a long-term monitoring optimization (LTM) application. Chadalavada and Datta (2008) developed optimal and efficient sampling locations for plume detection. Dhar and Datta (2007) developed a methodology based on the solution of optimization models for optimal design of groundwater quality monitoring networks. Reed et al. (2000) developed a methodology that combines a fate and transport model, plume interpolation, and a genetic algorithm to identify cost-effective sampling plans that accurately quantify the total mass of dissolved contaminant for

plume interpolation. Nunes et al. (2004) proposed three optimization models to select the best subset of stations from a large GMN. Zheng et al. (2005) did a similar work in concept to the work of Reed et al. (2000) using OK and inverse distance weighting (IDW).

Monitoring networks are integral to effective, efficient, and economical groundwater management. The objectives of designing optimal monitoring network may be different and vary as per site-specific conditions (Prakash and Datta 2015). To account for the transient and dynamic nature of groundwater flow and pollution transport, monitoring networks need to be designed and implemented sequentially with time. Sequential monitoring network design for transient transport process was considered by Grabow et al. (2000). Other previously reported works include sampling strategy in space and time using Kalman filter (Kollat et al. 2011) and integer programming for sequential monitoring network design (Dhar and Datta 2007; Mahar and Datta 1997). Kollat et al. (2008) developed a new multiobjective evolutionary algorithm (MOEA) to solve large, long-term groundwater monitoring (LTM) design problems. Time-varying dynamic monitoring network design methodology and its evaluation is reported in Chadalsavada and Datta (2008). Most of the optimization techniques for monitoring network design were aimed at detection of plume movement, with an implicit objective of minimizing the monitoring cost.

There exists a large body of literature dealing with design of optimal monitoring networks using heuristic optimization tools like genetic algorithm (GA) (Cieniawski et al. 1995; Wu et al. 2005; Yeh et al. 2006; Chadalavada et al. 2011) and Simulated Annealing (SA) (Prakash and Datta 2012, 2015) and genetic programming (GP) (Prakash and Datta 2013; Datta et al. 2013, 2014). MIP as the optimization algorithm together with chance constraints to define reliability of the monitoring network design was proposed by Datta and Dhiman (1996). Data interpolation techniques like geostatistical kriging have been used in groundwater monitoring network design (Yeh et al. 2006; Feng-guang et al. 2008; Chadalavada et al. 2011; Prakash and Datta 2012). Chadalavada and Datta (2007) proposed models for determining optimal and efficient sampling locations for contamination plume detection. Dhar and Datta (2010) developed an optimization-based solution for reducing the redundancy in a groundwater quality monitoring network. Bashi-Azghadi and Kerachian (2010) developed a new methodology for optimally locating monitoring wells using Monte Carlo technique. Other proposed methodologies include optimizing the groundwater monitoring network using probabilistic support vector machines (PSVMs) (Bashi-Azghadi et al. 2010), many-objective long-term groundwater monitoring (LTGM) network design and its trade-offs (Reed and Kollat 2012), and groundwater quality monitoring network using vulnerability mapping and geostatistics (Husam 2010). The main aim of all the above-mentioned methods was to reduce the cost due to redundancy in the monitoring, even then improved the detection of pollutants under conditions of uncertainty in a dynamic scenario. However, these methods did not address the need for an optimal sampling network that can improve the accuracy of pollutant source identification. A number of methodologies have been proposed using

different optimization algorithms for improving the source identification results as reported in Chadalavada et al. (2011).

Initial attempt to design an optimal sampling network that can improve the accuracy of pollutant source identification was addressed by Mahar and Datta (1997), Datta et al. (2009a, b), Prakash and Datta (2013, 2014). Datta et al. (2013) used GP-based monitoring factors for design of optimal monitoring network to improve the accuracy of pollutant source identification. Potential applicability of GP in groundwater problems was first discussed by Sreekanth and Datta (2012). These methodologies used trained GP models to calculate the impact factor of the sources on the candidate monitoring locations. However, these monitoring networks were not dynamic in design or implementation. The issue of sequentially designing monitoring networks to gather feedback information regarding the compliance with implemented management strategies for saltwater intrusion in coastal aquifers is discussed in Sreekanth and Datta (2014, 2015). Datta and Singh (2014) presented a methodology using kriging linked optimization model for contaminant monitoring network design by incorporating uncertainty.

The optimal monitoring network design methodology developed in this study utilizes a kriging-linked SA-based optimization model. It essentially utilizes a numerical flow and transport simulation model (MODFLOW and MT3DMS) to simulate the physical and geochemical processes. The next section describes the methodology adopted in developing the linked model.

2 Methodology

Kriging is a geostatistical estimation technique especially used in geologic and geophysical systems. The kriging linked SA model has three main components: (a) a groundwater flow and transport simulation, (b) a global mass estimation using Geostatistics, and (c) optimization using SA.

2.1 Groundwater Flow and Transport Simulation

The equation describing the transient, two-dimensional areal flow of groundwater through a nonhomogeneous, anisotropic, and saturated aquifer can be written in Cartesian tensor notation (Pinder and Bredehoeft 1968) as

$$\frac{\partial}{\partial x_i}\left(T_{ij}\frac{\partial h}{\partial x_j}\right) = S\frac{\partial h}{\partial t} + W; \quad i,j = 1,2 \tag{1}$$

where T_{ij} = transmissivity tensor (L^2T^{-1}) = $K_{ij}b$; K_{ij} = hydraulic conductivity tensor (LT^{-1}); b = saturated thickness of aquifer (L); h = hydraulic head (L); W = volume flux per unit area (LT^{-1}); and x_i, x_j = Cartesian coordinates (L). In this study, the flow model is simulated using MODFLOW (1996).

The partial differential equation describing the fate and transport of contaminants of species k in 3-D, transient groundwater flow systems can be written as follows (MT3DMS 1999):

$$\frac{\partial(\theta C^k)}{\partial t} = \frac{\partial}{\partial x_j}\left(\theta D_{ij}\frac{\partial C^k}{\partial x_j}\right) - \frac{\partial}{\partial x_i}\left(\theta v_i C^k\right) + q_s C_s^k + \Sigma R_n \tag{2}$$

where θ = porosity of the subsurface medium, dimensionless; C^k = dissolved concentration of species k, ML^{-3}; t = time, T; $x_{i,j}$ = distance along the respective Cartesian coordinate axis, L; D_{ij} = hydrodynamic dispersion coefficient tensor, L^2T^{-1}; v_i = seepage or linear pore water velocity, LT^{-1}; it is related to the specific discharge or Darcy flux through the relationship, $v_i = q_i/\theta$; q_s = volumetric flow rate per unit volume of aquifer representing fluid sources (positive) and sinks (negative), T^{-1}; C_s^k = concentration of the source or sink flux for species k, ML^{-3}; ΣR_n = chemical reaction term, $ML^{-3}T^{-1}$. MT3DMS (1999) has been used as contaminant transport simulation.

The contaminant plume simulated by flow and transport model represents the actual field measurements in the future time periods.

2.2 Geostatistics

ASCE task committee, (1990a, b) has defined the Geostatistics as a collection of techniques for making inferences about properties that vary in space. Geostatistics offers a collection of deterministic and statistical tools that provide understanding and modeling spatial variability. Ordinary kriging (OK) is the most commonly used variant of the simple Kriging (SK) algorithm. Kriging (SK or OK) has been performed to provide a "best" linear unbiased estimate (BLUE) for unsampled values (Deutsch and Journel 1998). Different semivariogram models are used in the kriging that are selected by performing a sensitivity analysis using different sets of parameters. For this study, spherical variogram model was selected and the package provided by GSLIB (1998) was modified to estimate the plume concentration at all the unsampled locations within the study area.

2.3 Optimization Algorithm: Simulated Annealing

The optimization algorithm used for the design of the monitoring network is based on SA. Annealing is the cooling process of molten metals. At the high temperature atoms with high energy move freely, and when the temperature is reduced get ordered and finally form crystals having minimum possible energy. This crystalline stage will not be achieved if the temperature is reduced at fast rate (Deb 2002).

Corana et al. (1987) presented a modified algorithm of Simulated Annealing. The SA parameters- temperature reduction factor, initial temperature, number of function evaluations for termination criteria are based on the sensitivity analysis as well as guidelines available in the literature (Deb 2002; Nunes et al. 2004).

3 Optimization Model Formulation

The objective is to determine the optimal set of the monitoring locations for which the normalized mass estimation error is minimum, while constraining the total number of monitoring wells (Singh 2008). The objective function is defined as

$$\text{Minimize}: \left| \left(\frac{M_{\text{cal}} - M_{\text{est}}^{k}}{M_{\text{cal}}} \right) \right| \tag{3}$$

Subject to

$$M_{\text{est}}^{k} = F\left\{ \underset{k}{KR}\left(C_{ij}^{s} \right) \right\} \quad \forall i,j \in k \tag{4}$$

$$C_{ij}^{s} = f(I, BC, S) \quad \forall i,j \in k \tag{5}$$

$$\sum_{m=1}^{M} W_m \leq W_M \quad \forall m \in M \tag{6}$$

where M_{cal} = total contaminant mass present in the study area; M_{est}^{k} = total estimated contaminant mass based on kth subset of candidate (optimal) monitoring locations; C_{ij}^{s} = simulated contaminant concentration at spatial location i, j belonging to the kth subset of the potential monitoring locations N at the end of the management period; $\underset{k}{KR}\left(C_{ij}^{s} \right)$ = spatially kriged concentrations at all nodes of the study area based on simulated concentrations at the kth subset potential monitoring locations; $F\left\{ \underset{k}{KR}\left(C_{ij}^{s} \right) \right\}$ = function of the spatially kriged concentrations C_{ij}^{s}; $f(I, BC, S)$ = predicted concentration at end of a management period obtained as solution of the simulation model based on defined initial conditions (I, observed at the implemented monitoring network at the end of last management period) at the beginning of the management period, boundary condition (BC) as specified, and source characteristics (S) if any known and specified. W_m = a binary decision variable, *1* indicates a monitoring well is selected at potential monitoring location, and *0* indicating otherwise; W_M = maximum permissible number of installed monitoring wells in a given management period; M = total number of potential

monitoring well locations; and m = maximum permissible number of potential locations in a design set.

In the first step of the proposed methodology, before applying the monitoring network design model to the field, it is essential to have a calibrated flow and transport model for the study area. The simulation model is used to simulate the contaminant scenario of the site from initial time t_0 to some expected time t_n. It is assumed to represent the future conditions to be monitored using the given initial conditions, boundary conditions, flow and transport parameters, contaminant source characteristics, and potential monitoring locations. The simulated concentration value at the potential monitoring locations are assumed to be known, and these known values of the contaminant concentrations are used to construct the contaminant plume using spatial extrapolation.

In the second step, the concentration values at the potential monitoring wells and their coordinates are used to randomly generate the specified number of the groundwater monitoring well-designed plans given as a maximum permissible number of wells in that particular analysis. In the third step, the optimization process starts with the given sets of the initialized parameters. At a given iteration, the SA algorithm chooses a subset of monitoring locations from the set of potential locations. This subset is then used as an input to the kriging model which estimates the mass of the contaminant based on the set provided by the SA. All these different sets are evaluated for the objective function and the algorithm terminates when the final function value at the current temperature differs from the current optimal function value by less than error tolerance of termination, and the optimal design set is evolved satisfying the imposed constraints.

4 Model Application

The size of the study area is 2800 m × 2700 m. The study area for illustrative application of the proposed monitoring network models are discretized into identical grids of 100 m × 100 m, as shown in Fig. 1.

The study area is assumed to have three continuous sources each with an injection mass flux of value 27.4 g/s. The injection rate of the contaminant sources is assumed to be 45 m^3/day which is approximately half liters per second. One hundred and eight observation wells are distributed in the area, for possibly collecting the concentration data in all management periods. The sources are running for all time periods. All other advection and dispersion parameters are described in Table 1.

The sensitivity analysis has been performed on SA model and Kriging model to find out the best sets of parameters utilized for the linked optimal model (Singh 2008).

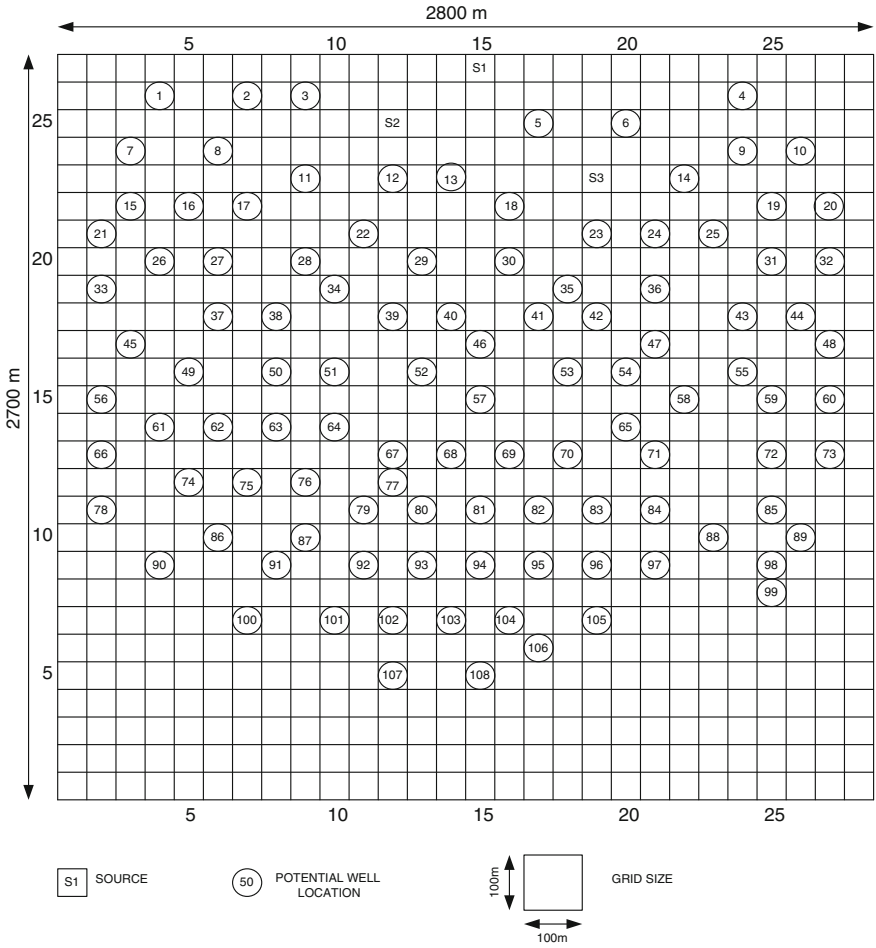

Fig. 1 Illustration of study area showing its physical parameters and potential monitoring well locations

4.1 Performance Evaluation

The performance of the proposed methodology is evaluated for an illustrated study for different contamination and management Scenarios.

4.1.1 Scenario 1

The study area as shown in Fig. 1 is considered as the aquifer which is unconfined, homogeneous, and isotropic in nature. All the flow and transport parameters are

Table 1 Flow and transport simulation model parameters

Parameters	Values
Porosity of the aquifer, θ	0.29
Grid size in x-direction, Δx	100 m
Number of nodes in x-direction, NX	28
Grid size in y-direction Δy	100 m
Number of nodes in y-direction, NY	27
Management period, t	365 days
Management time interval 1, Δt_1	90 days
Management time interval 2, Δt_2	275 days
Number of pumping wells	756
Number of injection wells	3
Mass flux of the injection wells	27.4 g/s
Injection rate, q_s	45 m^3/day
Storage coefficient	0.2
Longitudinal dispersivity, α_L	40 m
Ratio of the horizontal to the longitudinal dispersivity, α_{TH}/α_L	0.1

assumed to remain constant over time. The recharge rate, pumping rate, and boundary conditions change with time, contaminant sources are continuous, and mass flux rate is constant over time for all the sources. Three contaminant sources *S1*, *S2*, *S3,* and 108 potential monitoring well locations are considered. The monitoring network is designed for a management period of 1-year duration. Each 1-year management period is divided into two time intervals. Total mass estimate based on concentration values simulated at the end of the 1-year management period at the specified 108 potential monitoring locations constitute an input to the kriging linked SA model. The objective function which minimizes the mass estimation error is first utilized for solution of model. The numbers of monitoring wells are restricted to be 35, 45, and 55.

Table 2 gives the mass estimation error in percentage with the optimal number of wells as 35, 45, and 55. The error decreases with the increase in maximum permissible number of wells. This is expected, and it can be seen that the contaminant mass estimation error is not varying much from 45 to 55 wells.

Figures 2, 3, and 4 show the optimal locations of the 35, 45, and 55 wells respectively. The well locations chosen as solution of the design model are spatially spread over the study area the performance in terms of the mass estimation errors as shown in Table 1 can be judged as satisfactory. However, the order of the mass estimation errors decreases with 35 to 45 and 55 optimal monitoring locations.

Table 2 Number of wells and their corresponding mass estimation error in percentage in Scenario 1

No. of wells	Mass estimation error in %
35	0.2322
45	0.0021
55	0.0018

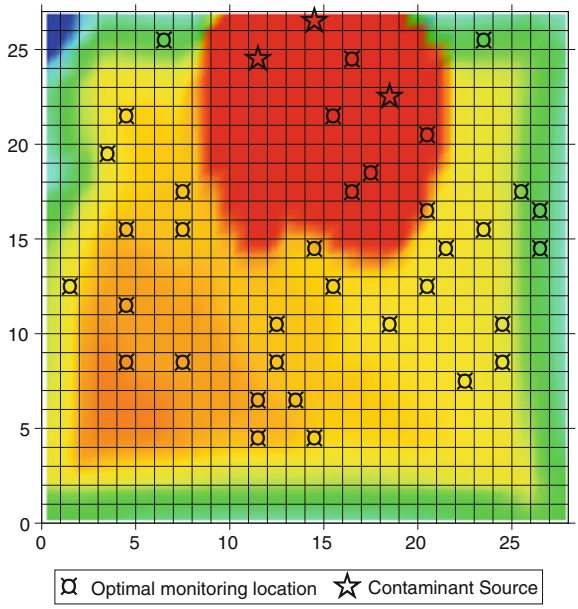

Fig. 2 Optimal location of monitoring wells for design set of 35 wells in Scenario 1

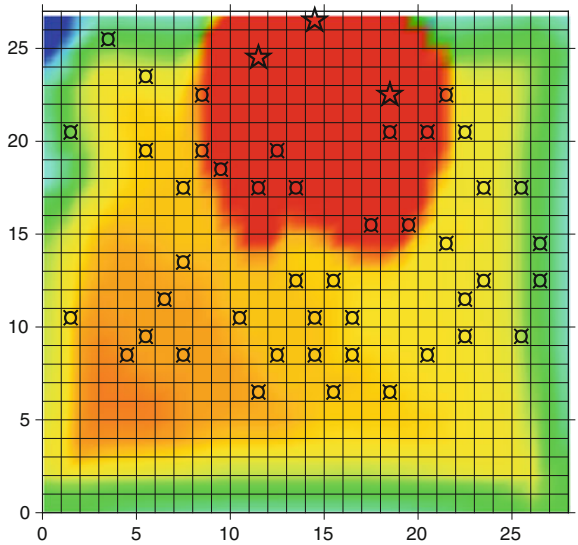

Fig. 3 Optimal location of monitoring wells for design set of 45 wells in Scenario 1

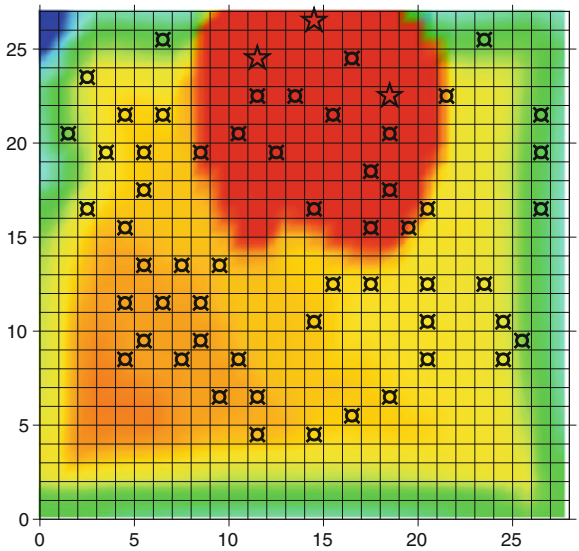

Fig. 4 Optimal location of monitoring wells for design set of 55 wells in Scenario 1

4.1.2 Scenario 2

All the conditions are same as in Scenario 1 except with the modifications in the contaminant source strength is considered uncertain. The mean contaminant source strengths are specified as 52608.0 mg/l. Uncertainties in terms of percentage are incorporated on the source concentration. Uncertainties in the source strength are incorporated by specifying a range of values for the source strengths. The range of upper bound and lower bound are shown in Table 3. These lower bound and upper bounds are utilized to generate different realizations of the source concentration using a uniform distribution. Ten realizations are generated from the uniform distribution for each of the sources. The simulation of the concentration plumes is performed for all 10 sets realizations. The mean concentration values at all the 108 potential monitoring locations obtained for these 10 realizations are utilized for spatial estimations of the concentration over the entire study area by kriging. This total contaminant mass estimate based on a number of realizations of the contaminant plume is used as input to the kriging linked SA model. Monitoring design model is solved to obtain the optimal monitoring network design.

Table 3 Range of contaminant source strengths for scenario 2

Uncertainty level (%)	Lower bound in mg/l	Upper bound in mg/l
10	47347.2	57868.8
20	42086.4	63129.6
30	36825.6	68390.4

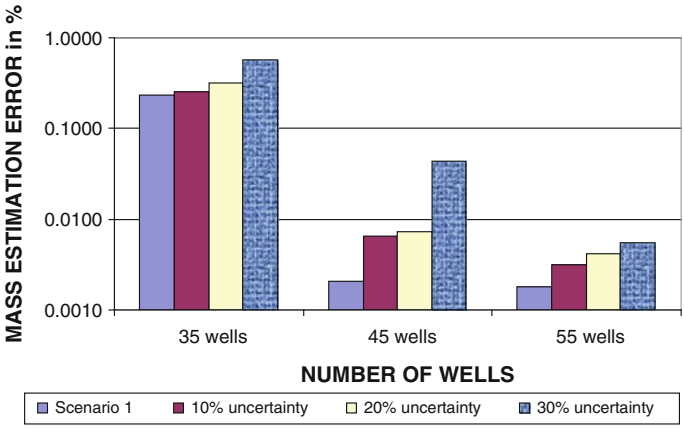

Fig. 5 Comparison of contaminant mass estimation error for scenario 1 and 2

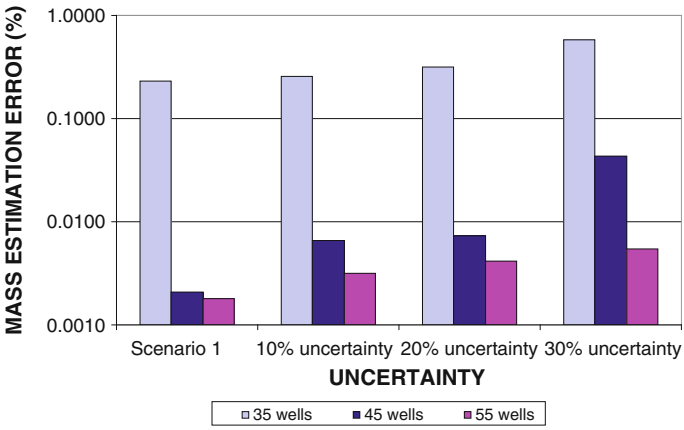

Fig. 6 Comparison of mass estimation errors with different levels of uncertainty

For each of the uncertainty levels, the optimal design set is obtained for 35 wells, 45 wells, and 55 maximum numbers of permissible wells. Figure 5 shows that for each set of wells, the mass estimation error increases with increase in the uncertainty level. The error values are compared with the values obtained for Scenario 1 without uncertainty. Figure 6 also shows that the mass estimation error decreases as the number of monitoring wells increases. These values are also given in Table 4.

Table 4 Mass estimation error in percentage for scenario 2

No. of wells	Uncertainty			
	Scenario 1	10 %	20 %	30 %
35	0.2322	0.2541	0.3139	0.5772
45	0.0021	0.0065	0.0073	0.0431
55	0.0018	0.0031	0.0042	0.0054

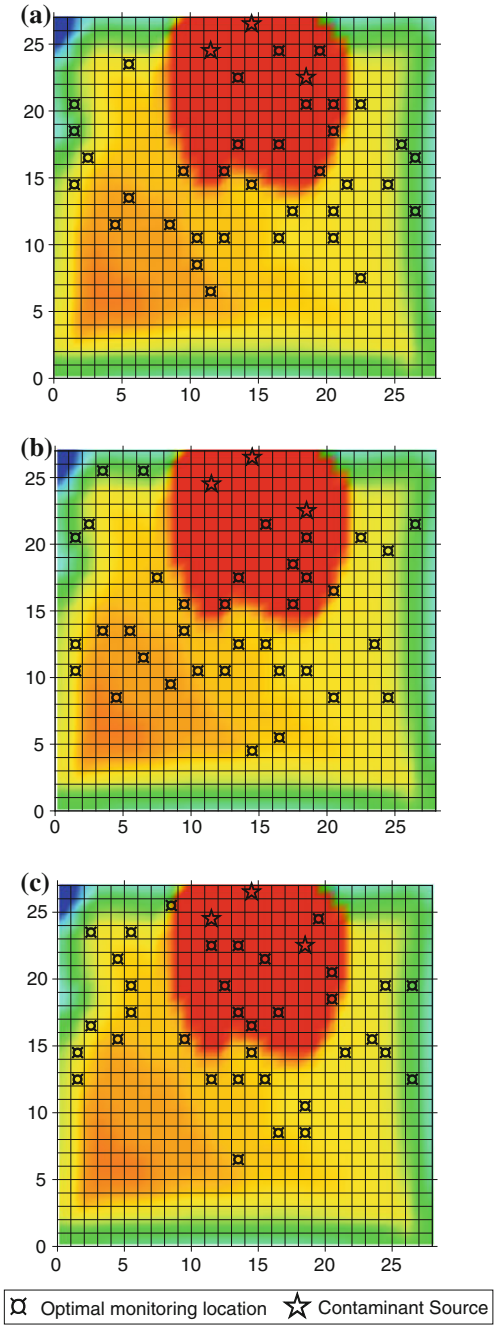

Fig. 7 Optimal monitoring locations for 35 monitoring wells with different uncertainties levels in scenario 2

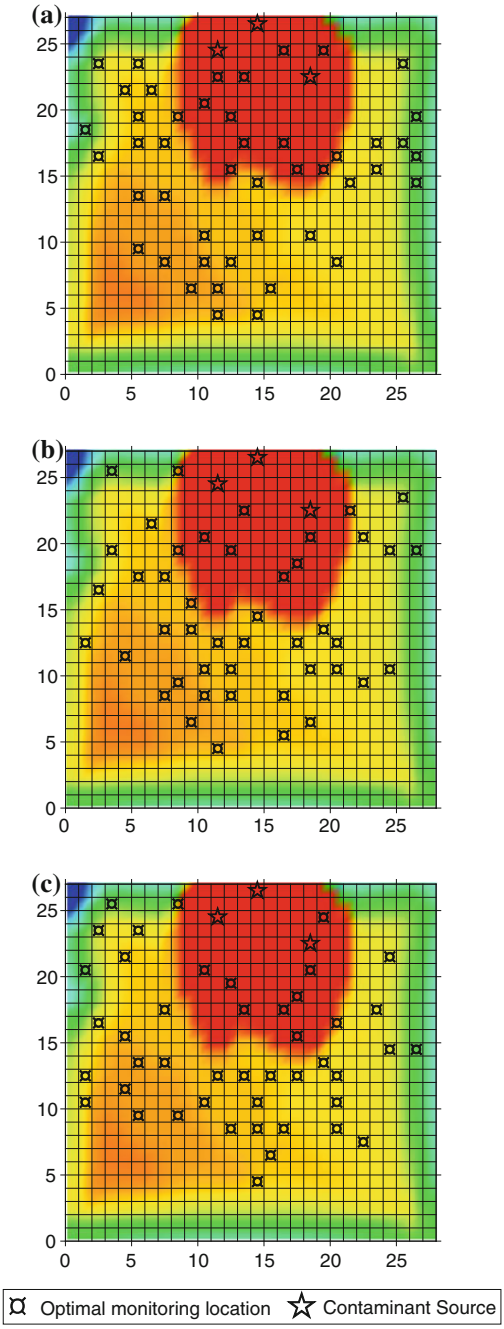

Fig. 8 Optimal design locations for 45 monitoring wells with different uncertainties in scenario 2

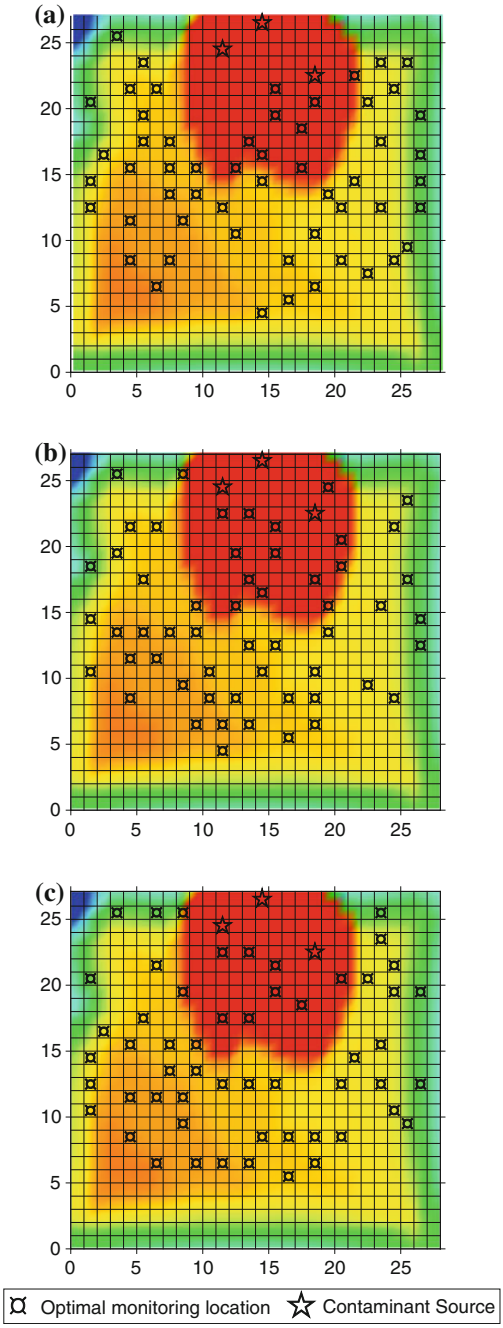

Fig. 9 Optimal design locations of 55 monitoring wells with different uncertainties in scenario 2

For the optimal design set of 35 wells the level of error is very high as shown in the Table 3. Some of the locations are common to all these designs. It can be noted from Table 3, that the mass estimation errors decrease with the increase in total number of monitoring wells, and it increases with the increase in the level of uncertainty in estimating the contaminant sources and therefore, the total mass of the contaminant. The optimal monitoring design locations for the different maximum permissible number of monitoring wells are shown in Figs. 7a–c, 8a–c, 9a–c. These figures show the variations in the designed networks for different permissible number of monitoring wells.

5 Conclusions

The developed optimization-based methodology for design of an optimal groundwater contamination monitoring network is solved for different flow and transport scenarios in an illustrative contaminated aquifer study area. The solution results are primarily useful for evaluation of the performance of the proposed models. These results appear to be consistent with intuitive solutions and therefore acceptable. However, much more rigorous performance evaluations may be necessary to establish the applicability of the proposed methodologies. The methodologies developed are applied to homogeneous unconfined aquifer for uniform parameters with transient contaminant transport process. Contamination sources are assumed to be conservative and continuous in nature. However, some of the following issues can be addressed in the future: (a) different types of contaminant sources varying in time and space may be considered for further evaluations; (b) the proposed methodology can be extended to non conservative and multi species contaminants in a 3-D system; (c) actual field data may be necessary before establishing the applicability of the developed methodologies. Also, the impact of measurement data reliability can be an important factor to consider.

References

ASCE Task Committee on Geostatistical Techniques (1990a) Review of geostatistics in geohydrology I: Basic concepts. J Hydraul Eng ASCE 116(5):612–632

ASCE Task Committee on Geostatistical Techniques (1990b) Review of geostatistics in geohydrology. II: Applications. J Hydraul Eng ASCE 116(5):633–658

Bashi-Azghadi BNS, Kerachian R, Lari MRB, Solouki K (2010) Characterizing an unknown pollution source in groundwater resources systems using PSVM and PNN. Expert Syst Appl 37 (2010):7154–7161