Optimal Identification of Ground-Water Pollution Sources

Hamid R. Ghafouri, Babak S. Darabi
Shahid Chamran University of Ahvaz, Ahvaz, I.R. Iran

Abstract: A new mathematical model for identifying pollution sources in aquifers is presented. The model utilizes Lagrange Constrained Optimization Method (LCOM) and is capable to inversely solve unsteady fluid flow in saturated, heterogeneous, anisotropic confined and/or unconfined aquifers. Throughout the presented model, complete advection-dispersion equation, including the adsorption as well as retardation of contaminant, is considered. The well-known finite element method is used to discretize and solve the governing equations. The model verification is implemented using a hypothetical example. Also, the applicability of the developed code is illustrated by the real field problem of Ramhormoz aquifer in southwestern Iran.

Introduction

The spread of hazardous pollutants in groundwater resources is a very important issue in many places. Finding the pollution source is the primary and crucial issue which would facilitate the planning of remedial measures. Although, the concentration of contamination could be measured using monitoring systems installed in any given site, it is often not known where the pollution comes from. This would usually necessitate the use of computer models in conjunction with measured field data. However, it is assumed in most predictive computer models concerning pollutant transport that the aquifer along with all its influencing parameters, including the characterization of heterogeneity and temporal variability as well as dispersion, are known. In addition, the characterizations of the source of pollution, i.e. location, magnitude and duration, are also known throughout these models. Obviously, the latter is not the case for our purpose. Lacking this information, can we determine where the pollution source is located? Many researchers have tried to answer this question by running advection-dispersion models backwards in time, with all inflows and outflows reversed. The method is closely related to the use of backward particle and front tracking, see for example; Schafer-Perini and Wilson[1991], Wilson and Liu[1995],. Their works would eventually yield two different types of probabilistic maps. The first type of map is called a travel time probability map. It describes how long it takes for water to move from some prescribed location in the aquifer to the well. It can be convoluted with estimated pollutant spatial concentrations to predict arriving pollutant concentrations. The second type of map, which is more relevant to the subject of the present paper, defines the origin of water observed or produced in the well. It gives the probability that pollution originating at some location in the aquifer will arrive at the well some specific time t later. This type of map, which is called the location probability map, is useful for identifying possible sources of past pollution and for monitoring design.

The problem was also challenged by others, amongst them the works of Skaggs and Kabala [1994], Sun [1994], Mahar et al.[2001] could be enumerated. Recently,

Singh et al. [2004] presented a solution for the problem using Artificial Neural Network (ANN) technique. A comprehensive review of the researches accomplished on the subject is presented by Atmadja [2001].

Unlike all previous models, throughout the model described herein, all computational grid points initially are considered as potential polluting sources. Also, another salient feature of the new model is the automatic elimination of nodal grid points possessing low concentration values. This is justified due to the fact that these are unlikely to coincide with pollution sources and the automatic removal of them from solution procedure will dramatically decrease the computational effort required.

Governing equations

The governing differential equation is obtained based on the mass balance of any particular solute in a control volume of porous media. The final unsteady form of the equation, including adsorption and other chemical reactions, is written as follows;

$$
\frac{\partial C}{\partial t} = \frac{\partial}{\partial x}\left(D_x \cdot \frac{\partial C}{\partial x}\right) + \frac{\partial}{\partial y}\left(D_y \cdot \frac{\partial C}{\partial y}\right)
$$
$$
+ \frac{\partial}{\partial z}\left(D_z \cdot \frac{\partial C}{\partial z}\right) - V_x \cdot \frac{\partial C}{\partial x} - V_y \cdot \frac{\partial C}{\partial y} \qquad (1)
$$
$$
- V_z \cdot \frac{\partial C}{\partial z} + CHEM
$$

in which C is the concentration of contamination at any given time and space, D the dispersion coefficient and V is the seepage velocity. The first three terms at RHS show the dispersion phenomena while the second three terms represent the advection of contaminant. Also $CHEM$ is consisting of the followings;

$$
- \frac{\rho_b}{n} \cdot \frac{\partial s}{\partial t} \qquad (2)
$$

which indicates the superficial forces of adsorption and ionization processes, and

$$
- \lambda \left(C + \frac{\rho_b}{n} s\right) \qquad (3)
$$

corresponds to the decay of pollutant. In the above equations ρ_b is the bulk soil density (ML^{-3}), λ is constant decay coefficient (T^{-1}), s is the density of absorbed substance on soil grains or colloids, and n is the soil porosity. Substituting these terms into equation (1) and simplifying them, the following equation is resulted in;

$$
\frac{\partial C}{\partial t} + \frac{\rho_b}{n}\frac{\partial s}{\partial t} = \frac{\partial}{\partial x}\left(D_x \cdot \frac{\partial C}{\partial x}\right) + \frac{\partial}{\partial y}\left(D_y \cdot \frac{\partial C}{\partial y}\right)
$$
$$
+ \frac{\partial}{\partial z}\left(D_z \cdot \frac{\partial C}{\partial z}\right) - V_x \cdot \frac{\partial C}{\partial x} - V_y \cdot \frac{\partial C}{\partial y} \qquad (4)
$$
$$
- V_z \cdot \frac{\partial C}{\partial z} - \lambda\left(C + \frac{\rho_b}{n} s\right)
$$

Assuming $s = k_d \cdot C$, k_d where is a distribution coefficient, applying the chain rule to the derivative of s and re-ordering the resulted relations, equation (4) may be written as follows;

$$
R\frac{\partial C}{\partial t} = \frac{\partial}{\partial x}\left(D_x \cdot \frac{\partial C}{\partial x}\right) + \frac{\partial}{\partial y}\left(D_y \cdot \frac{\partial C}{\partial y}\right)
$$
$$
+ \frac{\partial}{\partial z}\left(D_z \cdot \frac{\partial C}{\partial z}\right) - V_x \cdot \frac{\partial C}{\partial x} - V_y \cdot \frac{\partial C}{\partial y} \qquad (5)
$$
$$
- V_z \cdot \frac{\partial C}{\partial z} - R\lambda C
$$

The latter equation is three dimensional governing equation of advection-dispersion of contaminants in groundwater resources where $R = 1 + \frac{\rho_b \cdot k_d}{n}$ is the retardation factor.

Finite Element Discretisation

The analytical solution of equation (5) is usually unavailable for aquifers having irregular geometries and/or boundary conditions. Hence approximate numerical

solution of the equation is often sought in most cases. The method of finite elements is one of the most desirable methods frequently used for this purpose. In this method, using an imaginary computational grid, the solution domain is discretized into a number of sub-domains, or elements, inter-connected at nodal points. Then approximate solution of the sought unknown, i.e. the pollutant concentration (CL), is computed at nodal points. For any other given point, the concentration value is approximated as;

$$C(x,y,z,t) = \sum_{L=1}^{NNODE} C_L(t).N_L(x,y,z) \qquad (6)$$

in which $N_L(x,y,z)$ is an arbitrary approximating function, called shape function, of node L. Applying the finite element method to the equation (5) leads to a set of simultaneous algebraic equations as follows;

$$[A]\{C\}^{t+\Delta t} = [B]\{C\}^t + \{f\} \qquad (7)$$

where

$$[A] = [G] + [U] + [F] + \frac{1}{\Delta t}[P] \qquad (8)$$

$$[B] = \frac{1}{\Delta t}[P] \qquad (9)$$

Δt is the length of time interval, the so called time-step, $\{C\}^t$ is the vector of known concentrations at the beginning of any time-step and $\{C\}^{t+\Delta t}$ is the vector of sought unknown concentrations at the end of time-step. $[G]$, $[U]$, $[F]$ and $[P]$ are square matrices in which the number of rows and columns are equal to the number of nodal points in computational grid. The elements of these matrices are computed as followings;

$$G_{(L,i)}^e = \iint_e \left(D_x \frac{\partial N_i^e}{\partial x}.\frac{\partial N_L^e}{\partial x} + D_y \frac{\partial N_i^e}{\partial y}.\frac{\partial N_L^e}{\partial y} + D_z \frac{\partial N_i^e}{\partial z}.\frac{\partial N_L^e}{\partial z} \right) dxdydz \qquad (10)$$

$$P_{(L,i)}^e = \iint_e R.N_i^e N_L^e \ dxdydz \qquad (11)$$

$$U_{(L,i)}^e = -\iint_e \left(V_x^e \frac{\partial N_i^e}{\partial x}.N_L^e + V_y^e \frac{\partial N_i^e}{\partial y}.N_L^e + V_z^e \frac{\partial N_i^e}{\partial z}.N_L^e \right) dxdydz \qquad (12)$$

$$F_{(L,i)}^e = -\iint_e R.\lambda.N_i^e N_L^e \ dxdydz \qquad (13)$$

$\{f\}$ is the load vector calculated using the following boundary integral:

$$\{f\} = \oint_\Gamma (D_x \frac{\partial \hat{c}}{\partial x} n_x + D_y \frac{\partial \hat{c}}{\partial y} n_y + D_z \frac{\partial \hat{c}}{\partial z} n_z) N_L d\Gamma \qquad (14)$$

where $\hat{c}$ denotes the given concentration values at boundary nodes, n_i are directional cosines and Γ is the domain boundary.

If the source of pollution and its magnitude is known, the forward solution of equation (7) would result in the concentration of pollutant in various points of aquifers at any given time. This is implemented by initializing from a beginning time and marching step by step through time domain until the required elapsed time is reached. However, the implementation of the procedure would be impossible if the source of pollution is unknown. If this is the case, then equation (7) should be solved inversely to identify the source of pollution. The backward solution of the equation, however, requires a lot of trial and errors which makes it too cumbersome to solve, or even impossible in some cases. In such cases, the optimization techniques might be applied simultaneously to reduce the required number of trial attempts.

Backward tracking

Let's assume that the hydro-physical properties of an aquifer, such as dispersivity

D_x and D_y , retardation factor R , decay coefficient λ and etc. , are given based on the field measurements. Moreover, assume the observed concentration of pollutant is measured at a number of points in the aquifer (C_{obs}). Then the only remaining unknown parameters are the location, magnitude and duration of pollutant source (or sources) leaking contamination into the reservoir. In this case our aim is to find initial concentrations at various points of the reservoir (C_{ini}). Based on the initial concentration values, then the location and the magnitude of pollution sources would be identified. To achieve this, the governing set of equations is solved assuming trial initial concentration values. Using the aforementioned forward method, the concentration values resulted from the trial initial concentrations are then computed (C_{com}). If the initial concentration values are appropriately chosen, the computed and the measured final values must be identical. Otherwise, they would differ. The difference between the two sets then could be calculated as follows;

$$\overline{F} = \sum_{i=1}^{n} w_i \left(C_{obs} - C_{com} \right)_i^2 \qquad (15)$$

where i is the number of points in the computational grid having measured concentration values and w_i is the weight specified on the squared difference between estimated and observed concentrations. It is defined as [Mahar et al. (2001)];

$$w_i = \frac{1}{\left(C_{obs} + \eta \right)^2} \qquad (15\text{-}1)$$

The value of η can be taken as constant, sufficiently large such that errors at low concentrations do not dominate.

The entire procedure must be repeated using new modified initial concentration values until function $\overline{F}$ approaches its minimum.

The optimum initial values also would minimize the difference between C_{com} and C_{obs}. This would imply that C_{com} are identical to the real values and therefore the initial trial concentrations have been chosen correctly. The locations and the magnitude of the pollution sources will then be pin-pointed in terms of initial concentrations. Obviously, these are the points having the highest initial concentrations.

It is obvious from the above reasoning that the problem under consideration is in fact an optimization problem where the minimal of objective function is sought and the values of initial concentrations are main independent variables. On the other hand, in order to be physically meaningful, these variables must be positive and must not exceed a reasonable upper limit. This implies that a problem of constrained optimization is encountered. Next section therefore deals with optimization technique applied to the problem.

It is worth mentioning that appropriate assumption must be made for the duration of contaminant release. For the sake of simplicity, throughout the following proposed model a steady continuous source is assumed.

Constrained optimization technique

Numerous optimization techniques have been developed and successfully applied for problems arising in engineering. Amongst them, the optimizations methods utilizing the derivatives of objective function, usually indicate a faster convergence rate and are therefore more desirable for this class of problems. Hence, Augmented Lagrange Multiplier Method (ALMM), as described in Ref. [10] and [11], was chosen to solve the

problem, herein. The main motivation in ALMM is the reduction of dependency to the penalty function parameters. Upon using ALMM, the problem of constrained optimization will be reduced to an unconstrained problem, through which all the sought parameters are updated during optimization process. The procedure could be briefly described as follows;

Let's assume our objective is to find variable X so that function $\bar{F}(X)$ becomes minimal whereas the following constrains are also satisfied;

$$g_j(X) \le 0 \quad j = 1,2,3,...,m \qquad (16)$$

The first step is to substitute the inequality constrains by their equivalent equality forms. This is implemented by adding stack variables Z_j^2 to inequalities (16)

$$g_j(X) + Z_j^2 = 0 \quad j = 1,2,3,...,m \qquad (17)$$

It should be noted that regardless of Z_j value, Z_j^2 is positive. Without getting involved in detailed explanation, then we may combine both the objective function and the equality constraints into the following single relation (Ref. [11]);

$$A(X, \lambda_j, r_p) = \bar{F}(X) + \sum_{j=1}^{m} \left[\lambda_j \psi_j + r_p \psi_j^2 \right] \quad (18)$$

where

$$\psi_j = Max\left[g_j(X), \frac{-\lambda_j}{2r_p} \right] \qquad (19)$$

in which r_p and λ_k are optimization parameters. To solve the resulted unconstrained problem, the Variable Metric Method of optimization (VMM), as explained in Ref. [11], is used. The use of VMM will require optimal step for variation of optimization parameters. Finally, in order to determine the optimal step, the Quadratic

Interpolation Method (QIM), which uses the derivatives of objective function indirectly, is applied. The entire procedure is schematically illustrated in Fig. (1).

By utilizing the described procedure the location of pollution sources may eventually be found. However, in order to get a faster convergent rate and more precise results, an eliminative technique was also applied. It means that after each iteration, the resulted initial concentration values are evaluated for the entire domain and the nodal points having very little initial concentration values are deactivated automatically. Therefore the required time of solution for the following time-steps is dramatically reduced, as a large number of nodal points would be eliminated from the optimization procedure.

Model Verification

In this example a hypothetical homogeneous isotropic saturated confined aquifer with horizontal dimensions of 225 by 225 meters and a uniform thickness of 5 meters is assumed. Water flow is considered to be in x-direction only and assumed equal to 0.0001 m/s. The aquifer was modeled by a relatively coarse grid of 81 four-noded quadrilateral elements (Fig. 2). As illustrated in the figure, the sources of pollution have been located in nodes 28 and 84 where the pollution is assumed to be released into the aquifer continuously with concentration of 6 and 7 gr/lit, respectively. The retardation factor is equal to the unity and decay coefficient is assumed equal to zero. Distribution of pollution concentration at various points of aquifer is obtained by executing forward model for a 15-hour time period. In order to verify the developed program, these values are then inversely attributed to the nodal grid points as observed concentration values. The
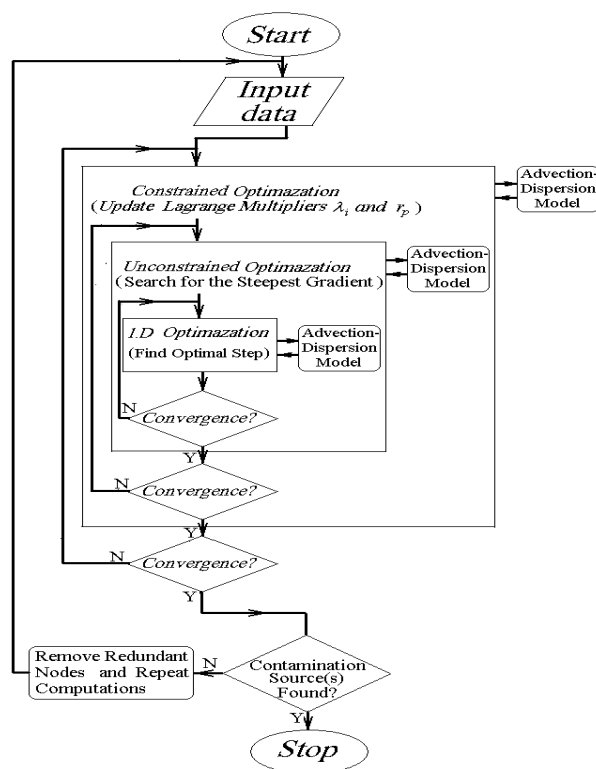
148

International Journal of Civil Engineerng. Vol. 5, No. 2, June 2007

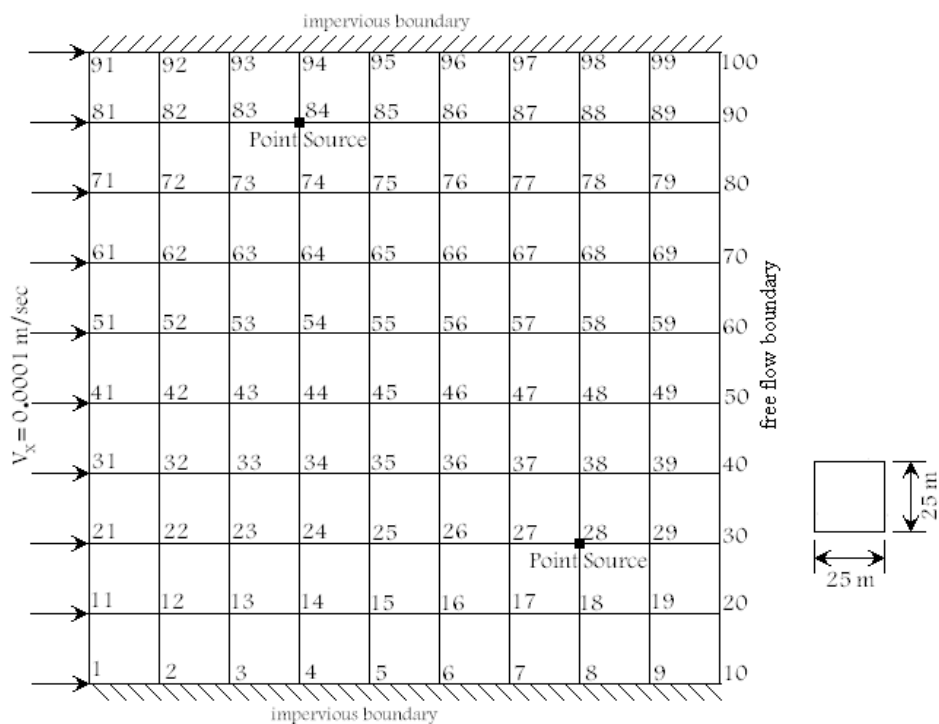**Fig.1.** The flowchart of computer code

Fig.2. The schematic representation of the problem.

program, when executed backward, should be able to identify the same sources of pollution on the basis of the given observed values. Figures (3a) to (3c) graphically indicate the results obtained by executing the developed code for the problem at three iteration levels. The dashed areas of the figures are the most probable locations of pollutants infiltration sources at each iteration level. As illustrated in the figures, these regions are gradually narrow down and eventually converge to very small zones. The converging points well coincide with nodes 28 and 84, i.e. the same initial selected pollutant origins. Therefore the applicability of the method is proved.

Model Application

As an example for practical use of the developed method described so far, it was also applied to a real field problem. The problem under consideration is the aquifer of Ramhormoz plain located at Khoozestan province southwest of Iran [Fig. (4)]. A brief description of the problem can be presented as follows: The area of the plain is about 483.5 km2 and its altitude varies between 75 to 250 meters from the mean sea level. The general slope of the plain is in northeast-southwest direction. The alluvial formations of plain are consisting of an unconfined aquifer having an approximate thickness of 100 meters. While the entire aquifer possesses an almost uniform storativity value of 0.026, it may be divided into 4 different regions in terms of transmissibility values, Table (1).

Total annual inflow into the aquifer is estimated 122.25E6 m^3. This involves 36.4E6 m^3 from 13 groundwater recharge regions mainly located at east, 2.9E6 m^3 from runoff at mountainous northeast areas,
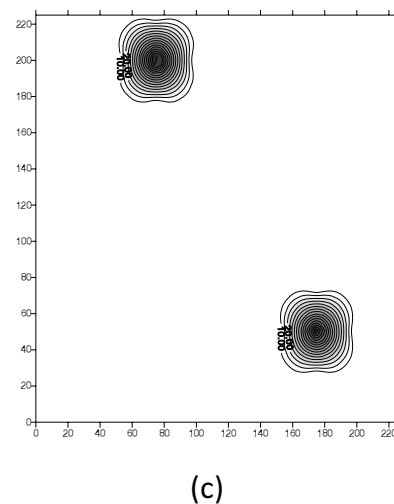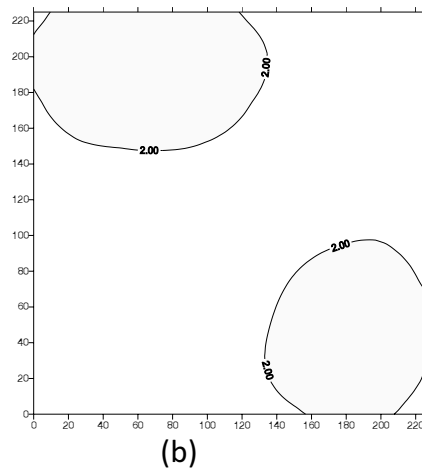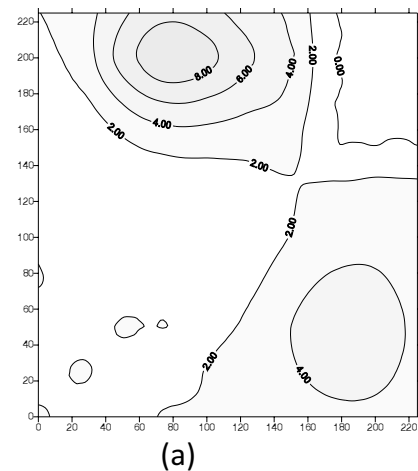

(a)

(b)

(c)

Fig.3. The highest initial concentration zones after (a) 3, (b) 8, (c) 12 iterations.

Table 1 Transmissibilities at various regions of the aquifer

region	Transmissibility (m^2/day)
1	250-300
2	600
3	1000-1500
4	2500-3000

Fig.4. General position of Ramhormoz at Southwestern IRAN.

50.4E6 m³ annual infiltrated precipitation, 2.15E6 m³ from Allah river direct infiltration and 14.2E6 m³ indirect infiltration resulted from agricultural activities and 16.2E6 m³ from other resources such as springs and wells, etc. .On the other hand, in addition to the natural drainage mechanisms, some 135 pumping well are producing water from the aquifer for various drinking, agricultural and industrial purposes. These are located all over the entire plain. Average estimated dispersivity of the aquifer is 10 m²/day whereas average porosity is 0.32. Based on the preliminary studies and trial runs, retardation factor and decay coefficient were assumed 2 and 0.3, respectively.

Measurements made at 37 observation wells during recent years - from 1997 on - had indicated that the aquifer was increasingly contaminated by some unknown pollution source(s). The contour map of Fig.(5) indicating the concentration of TDS1 at various points of the aquifer is constructed on the basis of measured concentration values. These values were applied to the model as observed concentration values. Although some origins might have been presumed for the contamination, none of them was definite and therefore it was a matter of long dispute among various responsible organizations.
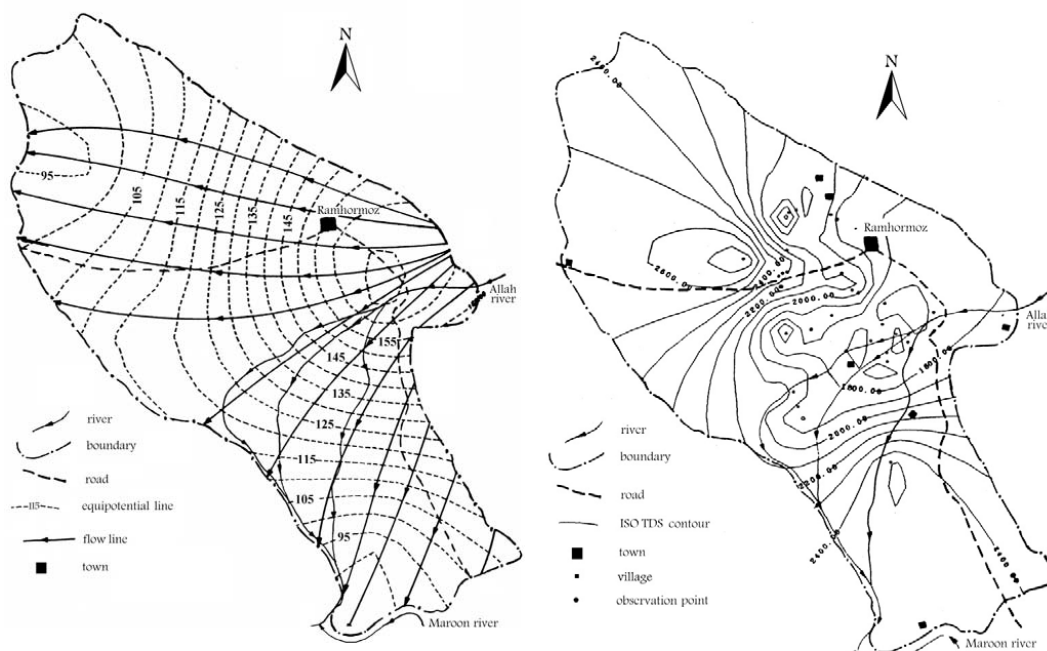
In order to identify the pollution source(s),

Fig.5. (left) Ramhormoz flownet and (right), iso- TDS map plotted on the basis of measured field data

the developed model was executed backward. In order to achieve this, the aquifer was modeled using 6-noded triangular elements of various sizes, as shown in Fig. (7). The computational grid was consisting of 234 elements and 132 nodes. Assuming a steady-state flow and utilizing initial hydraulic heads measured at various nodal points at March 1996, the flow net of the aquifer was constructed [Fig.(5-left)], based on which flow velocities at various points were computed and applied to the model. In order to inversely solve the advection-dispersion equation, a two-year period was divided into 24 one-month time steps. A trial uniform value of 500 mg/liter was considered as initial concentrations for all grid nodes.

Figures (7) through (9) indicate step-by-step progress of the process eventually leading to the precise identifying of contamination sources. It may be observed that the search

domain is more and more delimited, until it is converged to two relatively small zones. These are the most probable pollution sources as they have the highest initial concentrations. It is interesting to note that the center of the first zone almost coincide with the location of town Ramhormoz. The other zones, having fairly large areas, are places where agricultural activities are concentrated. The deduction is therefore that the contamination observed in the aquifer is originated mainly from domestic wastewater produced at Ramhormoz. This is well justifiable as Ramhormoz is lacking a proper sewerage collecting system or sewage treatment plant. Instead, the domestic wastewater is disposed underground using non-hygienic infiltration wells. Also, part of contamination could be attributed to the agricultural wastewater mainly produced at agriculture sites located nearby Ramhormoz. The magnitudes of the polluting sources are not identical, of course, as illustrated in Fig.9.
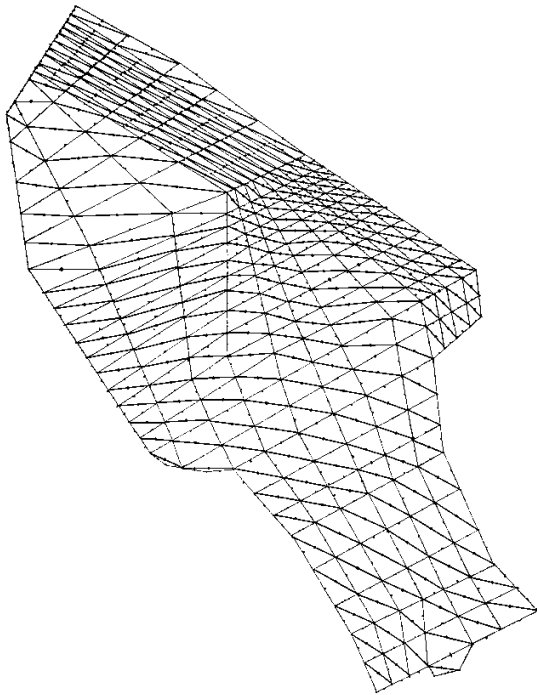
152

Fig.6. Computational Finite Elements grid.
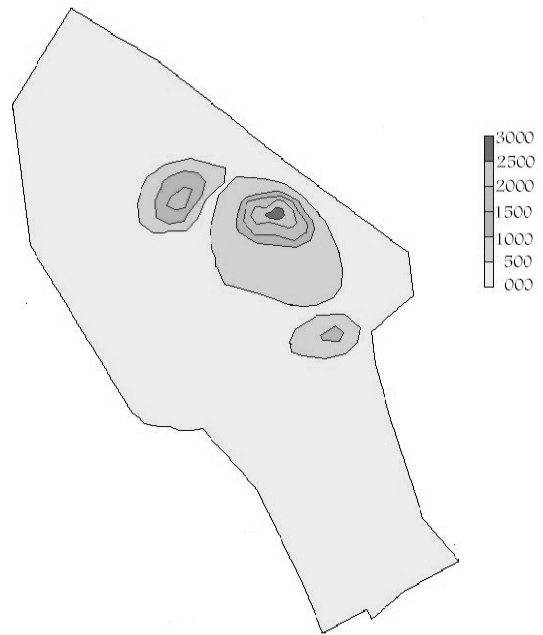

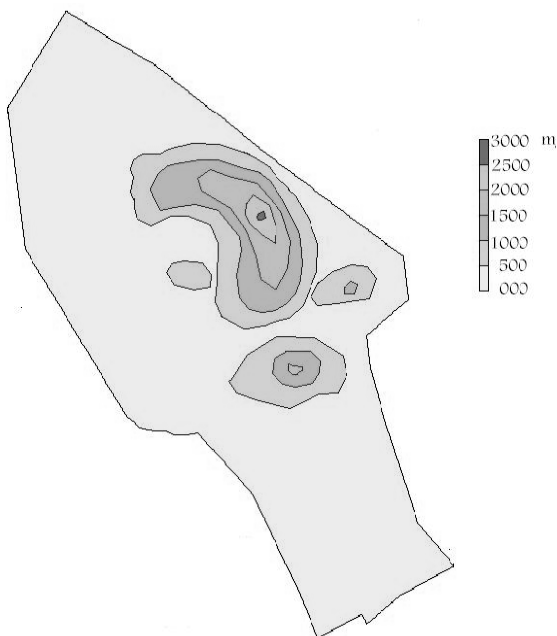**Fig.8.** The highest initial concentration zones, after 10 iterations.

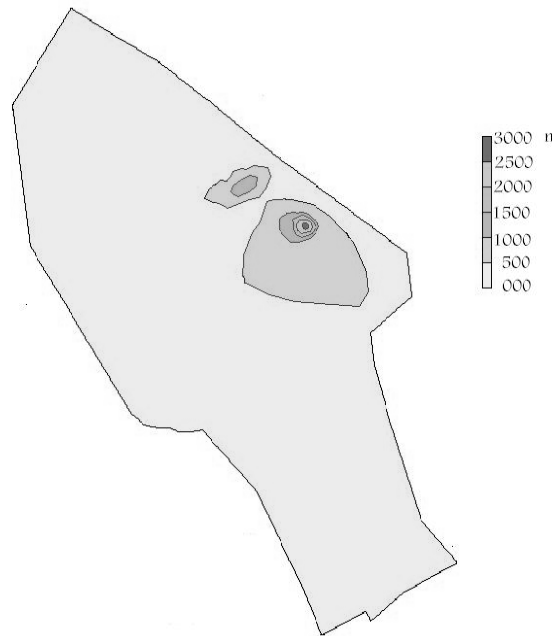**Fig.7.** The highest initial concentration zones, after 5 iterations.

Fig.9. The highest initial concentration zones, after 15 iterations.

Concluding remarks

The capabilities of optimization techniques as discussed here might be used for identifying pollution sources. Robustness of the method was practically indicated for two sample problems. In addition to precision associated with the method, it could be used for solving large scale problems because a fast rate of convergence could be achieved. Moreover, the methodology could be applied equivalently, when no initial guess whatsoever could be made for the location of probable pollution sources. This is because all grid nodes are assumed as potential sources. However, having an initial guess would facilitate the procedure as the search domain for optimized solution is delimited. This is important because otherwise, like many other optimization techniques, by converging to a local optimized solution a dummy source might be resulted. According to our experience, this may occur when the observed concentration values are distributed irregularly over the search domain and do not follow any particular pattern. Taking this into account, however, other global optimization techniques also might be used instead.

References

[1] Schafer-Perini, A.L., and J.L. Wilson, Efficient and Accurate Front Tracking for Two-Dimensional Groundwater Flow Models, Water Resources Research, v. 27, no. 7, 1471-1485, 1991.

[2] Wilson, J.L. and J. Liu, Backtracking to Find the Source of Pollution, Waste Management: from Risks to Reduction, R. Bahda et al., Eds.., ECM Press, Albuquerque, New Mexico, pp. 181-199, 1995.

[3] Skaggs, T.H. and Kabala, Z.H. Recovering the Release History of a Groundwater Contaminant, Water Resources Research 30(1),71-79, 1994.

[4] Sun, N.Z., Inverse Problem in Groundwater Modeling, Kluwer Dordrecht, The Netherland , 1994.

[5] Datta, B., Beggle, J. E., Kavaas, M., Land Orlab, G.T., Development of an Expert System Embedding Pattern Recognition Techniques for Pollution Source Identification, 1989.

[6] Mahar P.S. and Bithin Datta, "Optimal Identification of Groundwater Pollution Sources and Parameter Estimation", Journal of Water Resources Planning and Management, ASCE, 127(1), 20-29, 2001.

[7] Singh, R.; Datta, B., Groundwater Pollution Source Identification and Simultaneous Parameter Estimation Using Pattern Matching by Artificial Neural Network, Environmental Forensics, Vol. 5 Issue 3, p143, 11p-153, 2004.

[8] Atmadja, J.; Bagtzoglou, A.C. , State of the Art Report on Mathematical Methods for Groundwater Pollution Source Identification..., Environmental Forensics, Vol. 2 Issue 3, p205, 10p-214, 2001.

[9] Bagtzoglou, A.C., Dougherty, D.E, and Tompson, A.F.B., Application of Particle Methods to Reliable Identification of Groundwater Pollution Sources, Water Resources Management, 6,15-23,1992

[10] Murtagh, B.A., Saunders, M.A., A

Projected Lagrangian Algorithm and its Implementation for Sparse Nonlinear Constraints, 1982.

[11] Haftka, R. T., Gurdal, Z., Elements of Structural Optimization, 3rd Ed.,

Kluwer Academic Publishers, 1993.

[12] Pirhadi, A. , Study of Water Resources of Ramhormoz Plain (in Persian), Internal report, DEZAB Consulting Engineers Co., Ahvaz, Iran, 1998.