Singh, Raj Mohan; Srivastava, D.

Management of Pollution in Groundwater Systems Using Neuro-Fuzzy Model

Proceedings of ninth International Conference on Hydro-Science and Engineering (ICHE 2010), IIT Madras, Chennai, India.
2 - 5 August 2010

MANAGEMENT OF POLLUTION IN GROUNDWATER SYSTEMS USING NEURO-FUZZY MODEL

Dr. Raj Mohan Singh[1] and D. Srivastava[2]

Abstract: *The groundwater is a valuable natural resource. It is used through out the world as practical source of water for public supply agriculture and industry due to its quality, easy accessibility, reliability and relative low cost associated with its use. However, contamination of groundwater may prevent its use for drinking and other domestic and agricultural purposes. Modeling of subsurface flow and contaminant transport processes and its dynamics is basic necessity for dealing with groundwater management problems. In this study, neuro-fuzzy based methodology is developed to address uncertainty due to imprecision of flow and transport parameters in an aquifer. Developed methodology is demonstrated through example problems to identify the source fluxes responsible for groundwater pollution. A groundwater flow and contaminant transport simulation model is being utilized to simulate the pollution scenario in groundwater system. The simulated data is then utilized to develop the neuro-fuzzy based model. The neuro-fuzzy model, in fact, performs an inverse mapping to find the pollution source characteristics (location, magnitude and disposal periods) from observed pollution concentration in a specified number of observation wells. Performance of the model with increased number of simulation may improve the results.*

Keywords*: groundwater simulation; contaminant transport; pollution sources; neuro-fuzzy.*

INTRODUCTION

Good quality groundwater is one of the great natural resources. It is used through out the world as practical source of water for public supply agriculture and industry due to its quality, easy accessibility, reliability and relative low cost associated with its use. Contamination of groundwater poses serious threat to the environment. The contamination of aquifer not only threatens public health and the environment, it also involves large amounts of money in fines, lawsuits, and cleanup costs. Once groundwater is contaminated, it may be difficult and expensive to clean up.

---

[1] Assistant Professor, Department of Civil Engineering, Motilal Nehru National Institute of technology Allahabad, Allahabad 211004, India, Email: rajm.mnnit@gmail.com; rajm@mnnit.ac.in
[2] Ph.D. Scholar, Department of Civil Engineering, Motilal Nehru National Institute of technology Allahabad, Allahabad 211004, India, Email: divya@mnnit.ac.in

Sometimes it is impossible to clean it up to drinking water standards. Therefore, groundwater contamination is one of the most serious issues in groundwater management. Generally, we don't know about the source of pollution when pollution is detected in some supply or observation wells. The first step towards remediation is to identify the sources of pollution, which are responsible for the observed pollution scenario. Only then, the transport of pollution can be predicted, and a suitable remedial measure can be taken.

Gorelick *et al*., 1983 employed a groundwater transport simulation model incorporating linear programming and multiple regressions to estimate the source information. They defined the error as the difference between sampling concentration and simulated concentration. Then the linear programming method and multiple regression method were respectively used to minimize the sum of the absolute errors. Both methods could properly identify the source location, although the estimated release concentration was incorrect in the transient case. Hwang and Koerner (1983) employed a modified finite element model and limited monitoring well data to identify the pollution source by minimizing the sum of the squared errors between the sampling and simulated concentrations. Bagtzoglou et al. (1992) utilized particle methods to provide probabilistic estimates of source location and time history in heterogeneous site. Their study indicated that the simulation with a conditional conductivity field performs as well as the simulation with a perfectly known conductivity field. Mahar and Datta (2000, 2001) provided investigation to different types of source information estimation problems. In their study, the finite difference method was utilized to approximate a two-dimensional groundwater flow and the transport equation. They solved the optimization formulation of the source identification problem by non-linear programming. Their study successfully identified the source information for the flow in both steady and transient states. Aral *et al.* (2001) proposed the progressive genetic algorithm (PGA), in which the GA is combined with the groundwater simulation model, for the source identification problem. They demonstrated that the initial guess does not influence the identified solution. Mahinthakumar and Sayeed (2005) and Sayeed and Mahinthakumar (2005) employed hybrid genetic algorithm-local search (GA-LS) methods to solve the groundwater source identification problem. Their results indicated that the GA-LS methods were very effective to the groundwater source identification problem. Singh and Datta (2006) used genetic algorithm (GA) based simulation optimization approach for optimal identification of unknown groundwater pollution sources. A flow and transport simulation model is externally linked to the GA-based optimization model to simulate the physical processes involved. The main advantage of the proposed methodology is the external linking of the numerical simulation model with the optimization model. Mahinthakumar and Sayeed (2006) further compared four hybrid GA-LS optimization approaches for estimating the source information. The results showed that the release history recovery problem with the known potential source locations is much easier to solve than the source location identification problem.

Zou et al. (2007) proposed a neural network (NN)-embedded genetic algorithm (GA) for solving inverse water quality modeling problems to overcome the computational bottleneck of inverse modeling by replacing a water quality model with an efficient NN functional evaluator. Sun (2007) proposed robust geostatistical approach for contaminant source identification in a two-dimensional aquifer where the model uncertainty is caused by variability in hydraulic conductivity. Again Sun with his co-workers (2007) provided a constrained robust least squares

(CRLS) approach for contaminant release history identification. Such a strategy, however, requires detailed prior information on potential source locations. Prasad and Mathur (2007) developed a methodology wherein a genetic algorithm (GA) is used to find a global optimal solution to a groundwater flow and contaminant problem by incorporating an artificial neural network (ANN) to evaluate the objective function within the genetic algorithm. The study shows that an ANN-GA technique can be used to find the uncertainties in output parameters due to imprecision in input parameters. Ricciardi et al. (2009) presented one method for developing a pump-and-treat remediation design that satisfied both hydraulic head and concentration constraints while considering the uncertainty in the hydraulic conductivity. A search method that combines simulated annealing and a downhill simplex algorithm is used for determining the solution to the problem.

Results of most of the works show that groundwater flow and transport is sensitive to the variance of the uncertain flow and transport parameters, because the variance directly affects the source concentration in the optimization based model. To overcome this problem, fuzzy set analysis approach can be utilized to describe the uncertainty in aquifer parameters while providing a solution to the optimal groundwater management system. Fuzzy set theory provides an efficient mechanism for carrying out approximate reasoning processes when available information are uncertain, incomplete, imprecise, or vague. In this field, Coppla and his co-workers (2002) developed a fuzzy rule-based methodology for estimating monthly groundwater recharge. The fuzzy rule-based approach presented herein simplifies model input by using fewer and more easily quantifiable parameters. Zhang et al (2008) developed a methodology for incorporating probabilistic and fuzzy variables in one framework so as to solve a groundwater flow PDE (Probability density Equation) with uncertain parameter. In their work they showed how to apply the technique of hybrid uncertainty propagation through a groundwater model. Here, we proposed a methodology for identifying or characterizing unknown pollution sources based on fuzzy rule-base system based on fuzzy set theory. The main advantage of using Fuzzy set theory is that it provides an efficient mechanism for carrying out approximate reasoning processes when available information are uncertain, incomplete, imprecise, or vague.

GROUNDWATER FLOW AND TRANSPORT SIMULATION

Groundwater flow equation:

The proposed methodology is utilized in this study for groundwater systems under steady state flow conditions. The equation for the steady state two-dimensional areal flow of groundwater through a nonhomogenous anisotropic and saturated aquifer can be written in Cartesian tensor notation (Pinder and Bredeoeft 1968) as

$$\frac{\partial}{\partial x_i}\left(T_{ij}\frac{\partial h}{\partial x_j}\right)=W; \qquad \text{i,j=1,2} \qquad (1)$$

Where, T_{ij}= transmissivity tensor (L^2T^{-1})= $K_{ij}b$; K_{ij}=hydraulic conductivity tensor (LT^{-1}); and b= saturated thickness of aquifer (L); W =volume flux per unit area (positive sign for outflow and negative sign for inflow) (LT^{-1}); and x_i,x_j=Cartesian coordinates(L).

Contaminant transport equation:

The equation describing transient two-dimensional areal transport of a nonreactive, nonradioactive solute through a saturated aquifer through a saturated, rigid and nondeformable aquifer, in Cartesian notation, can be written (Bear 1972; Bredeoeft and Pinder 1973) as:

$$\frac{\partial(cb)}{\partial t} = \frac{\partial}{\partial x_i}(bD_{ij}\frac{\partial c}{\partial x_j}) - \frac{\partial}{\partial x_j}(bcv_i) - \frac{c'W}{\varepsilon}; \qquad \text{i,j=1,2} \qquad (2)$$

where t = time (T); c= concentration of the dissolved chemical species (ML^{-3}); D_{ij}= coefficient of hydrodynamic dispersion (second-order tensor) (L^2T^{-1}); c'= concentration of the dissolved chemical in a source or sink fluid (ML^{-3}); v_i= seepage velocity in the direction x_i(LT^{-1}); and ε = effective porosity of the aquifer (dimensionless). The groundwater flow and transport simulation model MOC (Method of characteristics) (Konikow and Bredehoeft, 1978) is being utilized to solve flow and transport equations.

NEURO-FUZZY BASED METHODOLOGY

As the complexity of the process being modeled increases, difficulty in developing dependable fuzzy rules and assigning adequate fuzzy membership functions increases. The drawback of fuzzy inference system (FIS) design is that the option of parameters in membership functions depends on the designer subjectively (Vadiee, 1993). This led to development of a class of hybrid architecture, a combination of fuzzy and artificial neural networks (ANN) methods called neuro-fuzzy, to improve the efficiency of the fuzzy rule base and to reduce the subjectivity in assigning parameters for the selected membership functions. The idea behind the fusion of the two powerful techniques is to use the learning ability of ANN's to implement and automate the fuzzy systems, which utilize the high level human-like reasoning capability. The neural networks' learning algorithms embedded in fuzzy systems provide a platform to adjust expert's knowledge, and automate parameters for the specified shape of the membership functions. This reduces the design time and cost. Basically, working with ANFIS means taking a FIS and tuning it with a neural network algorithm e.g. backpropagation algorithm (Rumelhaert et. al., 1986) or a hybrid algorithm (mixture of backpropagation and least squares) based on some collection of input-output data patterns. The neuro-fuzzy model tries to emulate dynamic, uncertain and non-linear systems by input-output data mapping and is effective especially when the underlying physical relationship is not fully understood.

Individual applications, basics etc. of either ANN or fuzzy is available in numerous papers and text (Haykin, 1994; Ross, 1997; etc.). Neuro-fuzzy model implemented in this study is a class of adaptive-network-based fuzzy inference system or adaptive neuro-fuzzy inference system (ANFIS) (Jang, 1993; Jang and Gulley, 1995). Some of the recent advancement of the neuro-fuzzy method is available in Nayak et.al 2005. The network structure and steps of learning of an

ANFIS is illustrated in Figure 2(a) and 1(b).

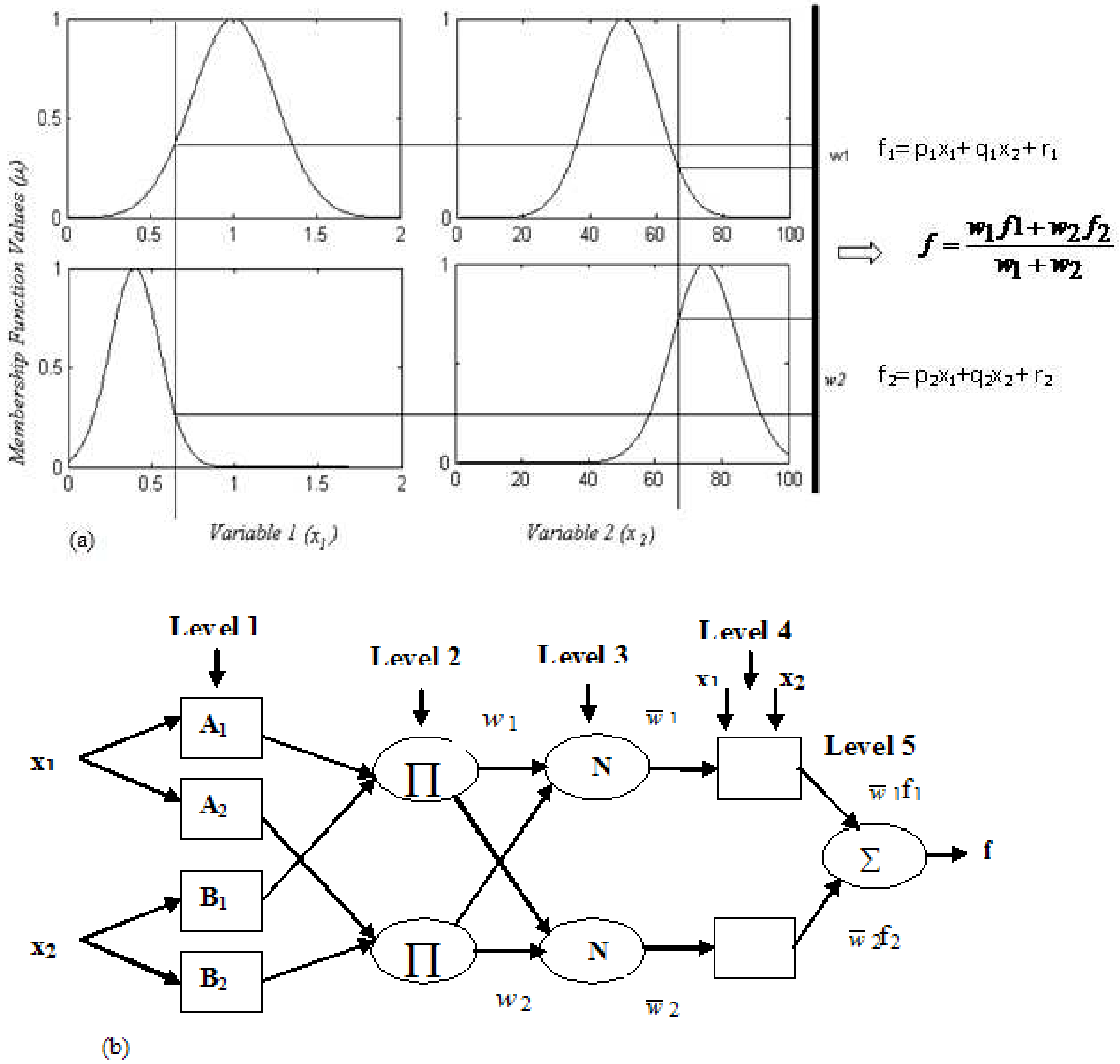

Fig. 2. Schematic representation of fuzzy and neuro-fuzzy models: (a) TSK fuzzy Inference System; (b) Architecture of equivalent ANFIS.

Development of Fuzzy Rule Based Methodology for Unknown Source Identification

ANFIS architecture which includes 5 layers or levels of connected nodes such as the fuzzification layer, product layer, normalization layer, defuzzification layer, and the total output layers. Layers and levels may be used intrechangably in this work. However, a subtle difference is meant by each representation. A layer is meant to represent only the connected nodes where as a level is meant to represent with incoming and outgoing signals (fuzzy operations) of a layer of connected nodes in Figure 2(b). The structure and functions of particular levels are can be found in the literature (Jang, 1993; Nayak et.al 2005).

For the formulation of contaminant transport problem in groundwater system, the process starts with random generation of source fluxes at potential source locations at different time interval. The Flow and Transport Simulation model MOC (Method of Characteristic) is utilized to simulate the contaminant concentration at observation location. The full breakthrough curve at observation location is then found out in terms of maximum, average, standard deviation, skewness and kurtosis for each of the source magnitudes. The MATLAB-NEURAL NETWORKS TOOL BOX software from MATLAB (2007) version was used to perform the necessary computations. Following steps were utilized to develop the neuro-fuzzy (ANFIS) based methodology to predict the contaminant transport in groundwater system:

1. Specification of hydro-geologic parameters, boundary conditions for the study area.
2. Random generation of source fluxes within a specified upper and lower value.
3. Simulation of pollution scenario for source concentrations for specified source magnitudes (randomly generated) and corresponding concentrations at observation locations.
4. Generation of pattern considering concentration characteristics as input and source magnitude as output.
5. Specification of number and shape of input membership functions for fuzzification of each of the inputs.
6. Training of ANFIS models for membership function parameters and linear parameter setting.
7. Testing the trained ANFIS models using data not used for training.
8. If results are satisfactory, design is complete.
9. If results are not satisfactory, repeat steps 5 to 8.

APPLICATION OF DEVELOPED METHODOLOGY

A hypothetical study area presented in Fig. 2 is selected for demonstrating performance evaluation of the developed methodology. A ten-year time domain divided into forty equal time steps is considered. The sources are assumed to release the pollutant in the aquifers during the first five years of the ten-year time domain. It is further assumed that the source releases the pollutants into the aquifer at a constant rate during one year.

The source magnitudes (500 in numbers) are generated randomly between a specified range for the potential location. Then, by using these randomly generated data, 100 simulations are performed taking each time 5 source fluxes that represent five years of individual responses respectively. The flow and transport simulation model, MOC, is utilized for these simulations. The Maximum, average, standard deviation, skewness and kurtosis of observed concentration are taken for each simulation. These five parameters are used as input for development of neuro-fuzzy base system to predict source fluxes from source during five years.

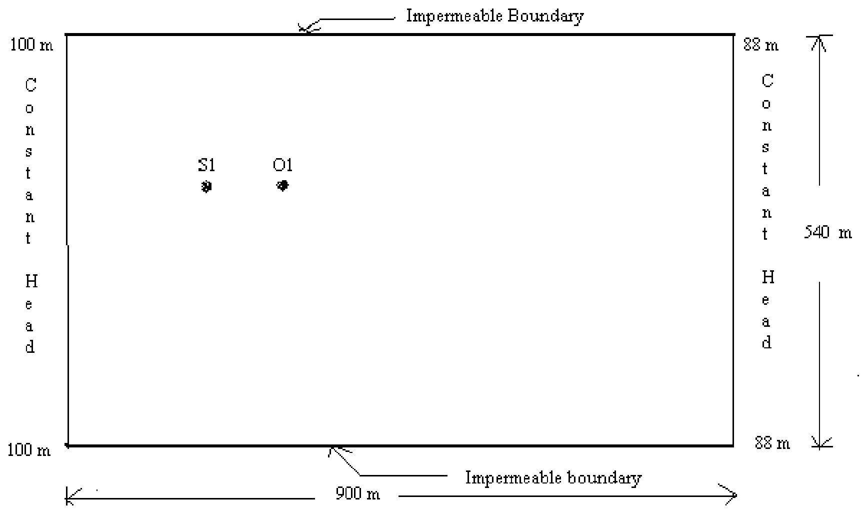

Fig.2. Hypothetical study area

PERFORMANCE CRITERIA

To evaluate the performance of the developed methodology, it is necessary to define the criteria by which performance is evaluated. To judge the predictive accuracy, the following statistical parameter is used for quantifying the errors.

Normalized Error (NE):

The NE, which is a measure of the methodology performance, is defined as:

$$\text{NE} = \frac{\sum (X_o - X_i)}{\sum X_o} \tag{3}$$

Model building: training and testing

The developed methodology is applied to identify the single source location utilizing observation from single observation well. The 70% of total data sets (100) is used for training and 30% for testing. The results obtained are presented in Table 1.

Table 1. Results for training & testing data

	Normalized Error (%)
Training data	22
Testing data	25

It can be seen from Table 1 that developed neuro-fuzzy based model shows comparatively better results for source identification in training as compare to testing.

COMPARISION WITH PREVIOUS STUDY

The results obtained by using fuzzy rule based approach showed NE=0.39 which is higher than the results obtained by using neuro-fuzzy based approach (NE=0.22). Thus developed model shows better performance than the optimization based approach in case of source identification.

The evaluation results show that developed neuro-fuzzy based methodology can be used to solve the complex problem of unknown groundwater pollution source identification. Limitations of this methodology in terms of large identification error are evident in the illustrative problems discussed here. The capability of neuro-fuzzy to provide approximate reasoning when available information are uncertain or imprecise are potentially suited to solve the unknown groundwater source identification problem. The performance of developed methodology does establish the potential applicability of the proposed methodology. The statistical analysis of results shows that developed model proves average capability of source identification of observed pollution.

CONCLUSIONS

The evaluation results show that developed neuro-fuzzy methodology can be used to solve the complex problem of unknown groundwater pollution source identification. Limitations of this methodology in terms of large identification error are evident in the illustrative problems discussed here. The capability of fuzzy set theory to provide approximate reasoning when available information are uncertain or imprecise are potentially suited to solve the unknown groundwater source identification problem. The performance of developed methodology does establish the potential applicability of the proposed methodology. The statistical analysis of results shows that developed model proves average capability of source identification of observed pollution.

REFERENCE

Aral, Mustafa M., Guan, J., and Maslia, M. L. (2001). "Identification of contaminant source location and release history in aquifers." *Jl. Of Hydrologic Engrg.,* ASCE, 6 (3), 225–234.

Atmadja, J and Bagtzoglou, A. C. (2001a). "Pollution source identification in heterogeneous porous media." *Water Resour. Res.,* 37(8), 2113 – 2125.

Atmadja, J and Bagtzoglou, A. C. (2001b). "State of the art report on mathematical methods to reliable of groundwater pollution source identification." *Environ. Forensics,* 2(3), 205-214.

Bagtzoglou, A. C., D. E. Dougherty, and A. F. B. Tompson. (1992). "Application of particle methods to reliable identification of groundwater pollution sources." *Water Resources Management*, 6, 15-23.

Bear, J. (1972). *Dynamics of fluids in porous media*, Dover Publication Inc., New York.

Bredehoeft, J.D., and Pinder, G. F. (1973). "Mass transport in flowing water." Water Resour

Res., 9(1); 194-210.

Coppola, E. A., Lucien Duckstein, L., and Davis, D. (2002). "Fuzzy Rule-based Methodology for estimating Monthly Groundwater Recharge in a Temperate Wtershed." *Jl. Of Hydrologic Engrg.,* ASCE, 7 (4), 326-335.

Gorelick, S. M. (1983). "A review of distributed parameter groundwater modeling methods." *Water Resour. Res.,* 19(2), 305-319.

Haykin, S. (1994). Neural Networks: A Comprehensive Foundations, Mac Millan, New York.

Hwang, J. C., and Koerner, R.M. (1983). "Groundwater pollution source identification from limited monitoring well data: part I, Theory and feasibility." *J. Hazard. Mater*., 8,105-109.

Jang, J.R. (1993). "Adaptive-network based fuzzy inference system." IEEE Trans. Syst. Man. Cybern., 23, 665-685.

Jang, J.R., and Gulley, N. (1995). Fuzzy Logic Tool Box For Use With MATLAB, The Math Works Inc., Natic Mass.

Konikow, L.F. and Bredehoeft, J.D. (1978). "Computer model of two-dimensional solute transport and dispersion in groundwater." *U. S. Geol. Surv. Tech. Water Resources Invest.book 7.*

Liu, C., and Ball, W. P. (1999). "Application of inverse methods to contaminant source identification from acuitard diffusion profiles at Dover AFB, Delaware." *Water Resour. Res.*, 35(7) 1975-1925.

Mahar, P.S., and Datta, B. (2000). "Identification of pollution sources in transient groundwater system." *Water Resource Management*, 14(6), 209 – 227.

Mahar, P.S., and Datta, B. (2001). "Optimal identification of ground-water pollution sources and parameter estimation." *Jl. Of Water Resources Plng. and Mgmt.*, ASCE, 127(1) 20– 29

Mahinthakumar, G., and M. Sayeed (2005). "Hybrid genetic algorithm-local search methods for Solving groundwater source identification inverse problem."*J. Water Resour. Plann. Manage*., 131(1), 45-57.

Mahinthakumar, G., and M. Sayeed (2006). "Reconstructing groundwater source release histories using hybrid optimization approaches" *Environ.Forensics*, 7(1), 20-29.

MATLAB. 2004. Language of technical computing. Fuzzy logic toolbox user's guide Mathworks, Natick, Mass.

Nayak, P.C., Sudheer, K.P., Rangan, D.M., and Ramasastri, K.S. (2005). "Short-term flood forecasting with a neurofuzzy model." Water Resour. Res., 41, W04004,1-16.

Neupauer, R. M., and J. L. Wilson (1999), "Adjoint method for obtaining backward-in-time location and travel time probabilities of a conservative groundwater contaminant." *Water Resour. Res*., 35(11), 3389-3398.

Neupauer, R. M., and J. L. Wilson (1999), "Adjoint method for obtaining backward-in-time location and travel time probabilities for a multi-dimensional groundwater system." *Water Resour. Res*., 37(6), 1657-1668.

Neupauer, R. M., and J. L. Wilson (2005). "Backward probability model using multiple observations of contamination to identify groundwater contamination sources at the Massachusetts Military Reservation." *Water Resour. Res*., 41.

Prasad, R. K. and Mathur, S. (2007). "Groundwater flow and contaminant transport simulation parameters." *J. of Irrigation. and Drainage Engg.,* 133(1), 61-70.

Pinder, G. F, and Bredehoeft, J. D. (1968). "Applicability of the digital computer for aquifer

evaluation." Wter Resour. Res., 4(5). 1069-1093.
Ross T.J. (1997). Fuzzy Logic with Engineering Applications. McGraw-Hill, Book Co., Singapore.
Rumelhart, D.E., Hinton, G.E., Williams, R.J. (1986). "Learning internal representation by error propagation." Parallel Distributed Processin. 1: 318-362, MIT Press, Cambridge, Mass.
Sayeed, M., and G. Mahinthakumar (2005) "Efficient parallel implementation of hybrid optimization approaches for solving groundwater inverse problems." *J. Comput. Civ. Eng.,* 19(4), 329-340.
Singh, R. M. and Datta, B. (2006). "Identification of groundwater pollution sources using GA-based linked simulation optimization model." *J. of hydrol. Engg.,* 11(2), 101-109.
Skaggs, T.H., and Kabala Z. H. (1994). "Recovering the release history of a groundwater contaminant." *Water Resour. Res.,* 30(1), 71–79.
Skaggs, T.H., and Kabala, Z. H. (1995). "Recovering the release history of a groundwater contaminant Plume: Method of quasi-reversibility." *Water Resour. Res.,* 31(11), 2669–2673.
Skaggs, T.H., and Kabala, Z. H. (1998). "Limitations in recovering the history of a groundwater contaminant plume." *J. Contam. Hydrol.,* 33, 347-359.
Snodgrass, M. F., and Kitanidis, P. K. (1997). "A geostatistical approach to contaminant source identification." *Water Resour. Res.,* 33(4), 537-546.
Sun, A. Y. (2007) "A robust geostatistical approach to contaminant source identification." *Water Resour. Res*., 43.
Vadiee, N. (1993). "Fuzzy rule-based expert systems –I and II." In M. Jamshidi, N. Vadiee and T. Ross (eds.), Fuzzy logic and control: software and hardware applications, Prentice Hall, Englewood Cliffs, N.J., Chapters 4 and 5.
Wagner, J. M., Shamir, U., and Nemati, H. R. (1992). "Groundwater quality management under certainty: Stochastic programming approaches and the value of information." *Water Resour. Res.,* 28(5), 1233-1246.
Woodbury, A. D. and Ulrych, T. J. (1996). "Minimum relative entropy inversion: the release history of a groundwater contaminant." *Water Resour. Res.,* 32(9), 2671-2681.
Woodbury, A. D., Sudicky, E., Ulrych, T. J and Ludwig, R. (1998). "Three dimensional plume source reconstruction using minimum relative entropy inversion." *J. Contam. Hydrol.,* 32, 131-158.
Zou, R., W. Lung, and J. Wu (2007). "An adaptive neural network embedded genetic algorithm approach for inverse water quality modeling" *Water Resour. Res*., 43.