LEAST COST SEARCH ALGORITHM FOR THE IDENTIFICATION OF A DNAPL SOURCE

Z. DOKOU and G. F. PINDER

Research Center for Groundwater Remediation Design, University of Vermont, Department of Civil and Environmental Engineering, Votey Building, , Burlington, VT, 05401, USA.

ABSTRACT

The goal of this work is to identify the source of DNAPL contamination using an optimal search algorithm which exploits the observation that a plume emanating from a DNAPL source is more easily defined that the source itself. In the proposed algorithm, target locations of the possible source are identified and given initial weights using an approach called information fusion. Given the initial identification of the sources the overall strategy uses stochastic groundwater flow and transport modeling where the hydraulic conductivity is assumed known with uncertainty (Monte Carlo approach). The hydraulic conductivity realizations are obtained using the Latin hypercube sampling strategy. The algorithm defines how to achieve an acceptable level of source-location accuracy with the least possible number of water quality samples. Each new concentration sample selected is the one that reduces the total concentration uncertainty the most. After each sample is taken, the plume is updated using a Kalman Filter. The plumes emanating from each individual source are calculated using the Monte Carlo approach and are compared with the updated plume. The scores obtained from this comparison are used as input weights for each individual source, and the above steps are repeated until the optimal source location is found.

1. INTRODUCTION

A DNAPL source location is generally too small and filamentous to identify via borings or geophysical methods. On the other hand, the plume emanating from a DNAPL source is typically quite large and consequently easily discovered, although identification of its extent and its concentration topology may, depending upon the nature of the groundwater flow field, require the collection of considerable field data. The goal of this work is to identify the source of DNAPL contamination using a search algorithm which exploits the above observation. Water quality, lithological and permeability data are the primary field data used in this work.

In this presentation the following assumptions are made: i) a groundwater plume has been identified and a preliminary field investigation has been conducted, ii) there is reason to believe that the plume is generated by at least one of the suspected potential DNAPL sources, iii) enough hydrological information on the site exists to construct a groundwater flow and transport model, assuming the hydraulic conductivity is known with uncertainty, iv) the primary source of uncertainty in the transport equation is the velocity which, in turn, is due to the uncertainty and heterogeneity in the hydraulic conductivity (the porosity, dispersivity, retardation and chemical reaction are assumed to be deterministic), and v) the proximity of the

DNAPL source can be identified with a DNAPL species concentration above a pre-determined concentration (the dissolution concept).

2. BACKGROUND AND METHODOLOGY

The approach used in this paper is different from the commonly encountered inverse modeling approach (see Skaggs and Kabala, 1994; Neupauer and Wilson, 1999, 2001; Wagner, 2002; 2003, Baun and Batzoglou, 2004; Neupauer, 2004; Michalak and Kitanidis, 2004). In contrast, the methodology employed in this work can be summarized as follows:

First, we have to assemble and interpret all available geohydrological information needed to construct a groundwater flow and transport model of the site that utilizes a random field of hydraulic conductivity and an uncertain source location.

To model this system a Monte-Carlo technique is used wherein realizations of the random hydraulic conductivity field are required. These realizations are generated using the Latin-hypercube sampling strategy which is a stratified sampling technique where the assumed probability density function is divided into a number of non-overlapping intervals. Each interval represents equal probability. Samples are taken, one from each area, and are permuted in a way such that the correlation of the field is accurately represented. The use of Lhs only requires the prior knowledge of the statistics of the field (Zhang and Pinder, 2000).

A number of potential DNAPL source locations are identified and each is associated with an initial weight that reflects our confidence that it is the true source location. These initial weights are determined using an approach called information fusion. In this approach each possible source location is described by an 3-dimensional vector, whose coordinates are values of identifying features of the source, such as its proximity to a manufacturing facility (A) and a waste dump (B), and the distance of the water table from the ground surface (C). For each feature a membership function capturing the meaning of “near” is provided by an expert and it is used to obtain the membership degree of each feature value for a particular site. In addition, the expert provides monotone measures which contain all the information about the importance of each individual feature and all groups of features for identifying the true source. Using the Choquet integral all the individual scores are combined and a global degree of confidence of the statement ‘source location *i* belongs to the group of true source locations’ is assigned to each possible source location. These weights reflect the number of times each source will be considered when calculating the concentration realizations. More information on the information fusion technique and the Choquet integral can be found in Dubois and Prade, 2004, Klir et al.,1997, Grabisch, 1996.

After the hydraulic conductivity realizations are obtained, the flow and transport equations are solved numerically using each realization to produce the contaminant concentration realizations (Monte Carlo approach). From the concentration realizations the concentration statistics (means and variances) are calculated.

At this point the water quality data are incorporated into the search strategy. The Kalman filter is used to determine where to collect a concentration sample. An important concept in the Kalman filter is that although we do not know the concentration value at points where water quality samples have not been taken, we do know that the uncertainty at any point where a sample is taken reduces to the sampling error. The Kalman filter uses this information to determine the impact of sampling at a particular location to the overall uncertainty of the concentration field. The optimal sampling location is the one that reduces the overall uncertainty of the field the most. After a sample is taken the Kalman filter is used again to update the concentration mean and variance-covariance matrix using the real data. A detailed description of the above use of the Kalman filter is provided by Herrera, 1998.

We return now to the source location alternatives. A concentration random field is produced for each different source alternative using the Monte Carlo approach and the field statistics are calculated. Each individual source location plume is compared to the updated plume using the following method. The plumes are represented as fuzzy sets with membership functions defined as normalized concentration values. Several α-cuts of the fuzzy sets are considered. An α-cut is a crisp set that contains all the elements of a fuzzy set whose membership degrees are greater or equal to the specified value of α. In this work we used 10 α-cuts (α=0.1,0.2,…,1). Each α-cut for the updated plume was compared with the corresponding α-cut of the individual plume and the number of nodes that were common in the 2 α-cuts was recorded. The global degree of similarity between the two plumes was obtained by weighting the number of nodes present in the two α-cuts by the α value itself. This way we emphasize the intersection of the two plumes and we weight more the higher concentration values. The degree of similarity between each individual source location plume and the updated plume is normalized and assigned as a new weight to each potential source location.

The procedure of obtaining the concentration field is followed using the new weights and a second sample is taken (after the mean and variance-covariance matrix of the plume has been updated with the first sample using the Kalman Filter) at a location that will reduce the new total uncertainty the most. The process is repeated until convergence on an optimal location is achieved.

3. RESULTS

The purpose of this section is to demonstrate how effectively the DNAPL source search strategy finds the true source location by applying the methodology to two synthetic example problems. The advantage of using synthetic models is that the true source location is known, which allows for the testing of the search strategy. For both problems, one of the potential source locations was selected as the true one and a contaminant plume was generated using one realization of the hydraulic conductivity field. When a sample was needed, the calculated concentration obtained using this single-realization model at the location of the proposed sampling well, was used as the surrogate for the concentration that would be measured in the field.

3.1 Test problem 1.

The hypothetical aquifer system used in this example is shown in Figure 1a. The aquifer is 1000m long and 500m wide, with constant head boundaries along the left side (h=1m) and right (h=0m) side. The mean hydraulic conductivity is 5m/d. There are 6 potential source locations, shown in Figure 1a, and the true one is assumed to be number 1. The number of potential sampling wells is 70 and they are shown in Figure 1b. Figure 2 shows the true plume generated by a single hydraulic conductivity realization assuming that the true source location is number 1.

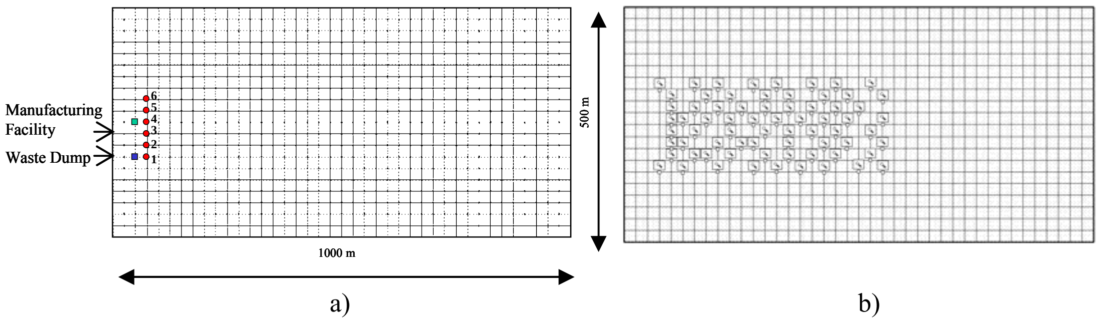

FIGURE 1. a) Synthetic aquifer for example 1 b) Potential water quality sampling locations

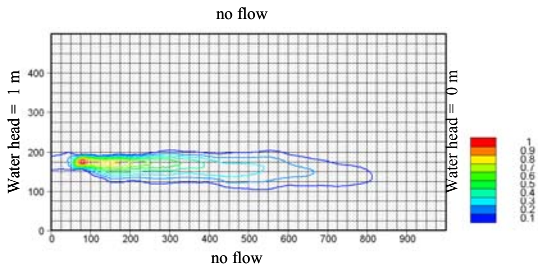

FIGURE 2. True plume generated by a single realization of hydraulic conductivity

Figures 3 and 4 show the weights and updated plumes that are obtained at the first iteration and after 1, 4 and 7 samples are taken. We are not showing all the intermediate steps due to lack of space. The initial weights shown in Figure 4 are the ones obtained using the information fusion technique (Choquet integral). We can see that the search algorithm is able to identify the true source location after taking 7 concentration samples. The location where a new concentration sample is taken at the current iteration is shown by a red dot. Concentration samples taken at previous steps are shown by black dots.

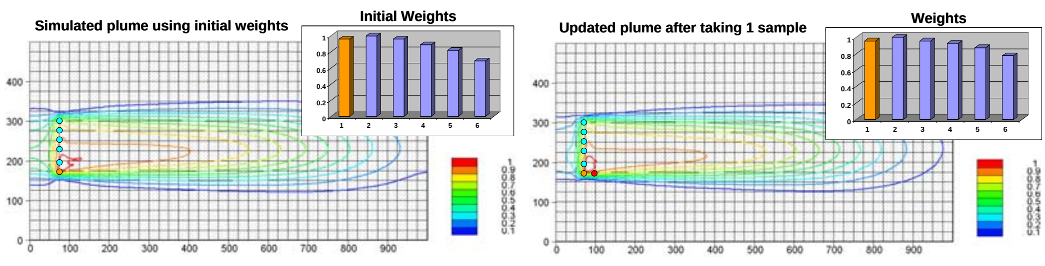

FIGURE 3. Simulated and updated plume and initial and obtained source location weights at the first iteration and after taking 4 concentration samples

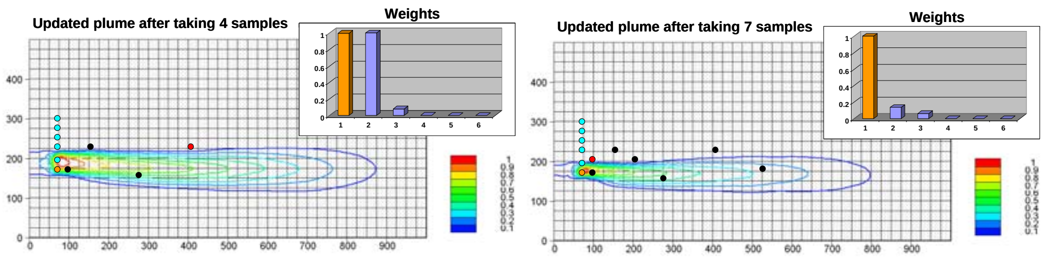

FIGURE 4. Updated plumes and obtained weights after taking 4 and 7 concentration samples

3.2 Test problem 2.

The hypothetical aquifer system used in this example is similar to the one used in the previous example, the only difference is that in this case there is a pumping well located in the upper middle part of the aquifer, as shown in Figure 5. Figure 5 also shows the true plume generated by a single hydraulic conductivity realization assuming that the true source location is number 1.

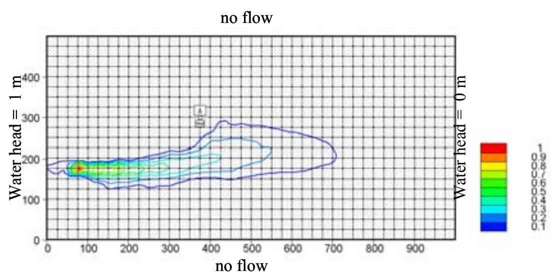

FIGURE 5. True plume generated by a single realization of hydraulic conductivity and location of pumping well

Figures 6 and 7 show the weights and updated plumes that are obtained at the first iteration and after 3, 7 and 10 samples are taken. We can see that the search algorithm is able to identify the true source location after taking 10 concentration samples. There are more samples needed in order to find the true source in the case of the well example. This is expected since the individual plume geometries are affected by the well and tend to look more similar. Thus, the algorithm needs more steps in order to reach convergence. In addition to that, the weight of source number 2 is not reduced as much as in the first example.

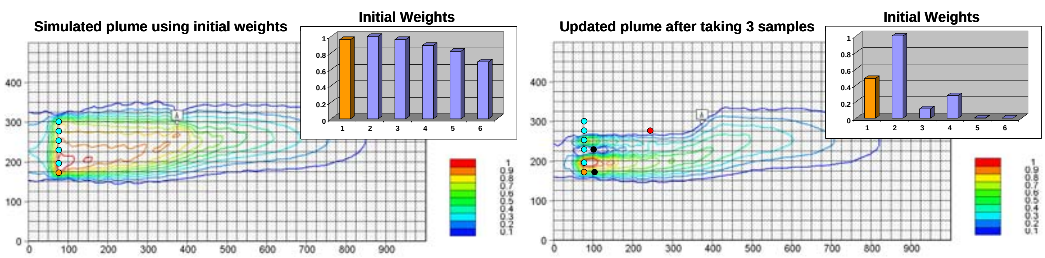

FIGURE 6. Simulated and updated plume and initial and obtained source location weights at the first iteration and after taking 3 concentration samples

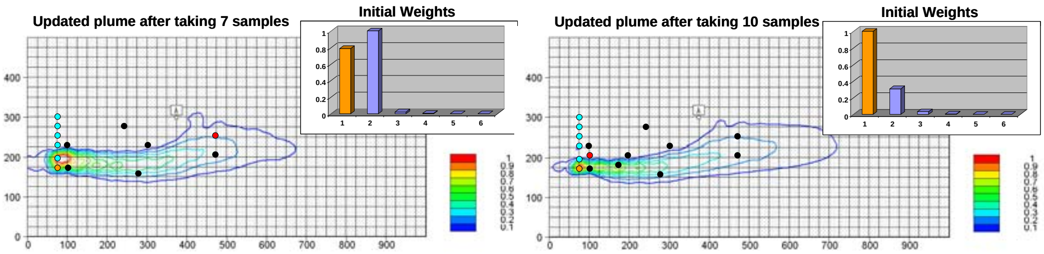

FIGURE 7. Updated plumes and obtained weights after taking 7 and 10 concentration samples

4. CONCLUSION

The search algorithm converges to the true source in both examples outlined above. There are 7 concentration samples needed in the first example and 10 samples needed in the second example in order for the algorithm to converge. There are 3 more samples needed in order to find the true source in the case of the well example. This is not surprising if we consider the fact that the individual plume geometries are affected by the well and tend to look more similar. Thus, the algorithm needs more steps in order to reach convergence. We should also note that the weight of source number 2 is not reduced as much as in the first example. That can also be explained by the fact that the individual plumes tend to look more similar in the well case.

ACKNOWLEDGMENTS

This material is based upon work supported by the Strategic Environmental Research and Development Program (SERDP) under Contract No. DACA72-03-C-0015.

Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Strategic Environmental Research and Development Program (SERDP).

REFERENCES

Baun, S. A., and A. C. Batzoglou (2004), A computationally attractive approach for near real-time contamination source identification, *Proceedings Computational Methods in Water Resources*, Chapel Hill, USA, 1263-1271.

Dubois, D., and H. Prade (2004), On the use of aggregation operations in information fusion process, *Fuzzy Sets and Systems, 142*, 143-161.

Grabisch, M. (1996), The application of fuzzy integrals in multicriteria decision making, *European J. of Oper. Res.*, *89*(3), 445—456.

Herrera de Olivares, G. (1998), Cost effective groundwater quality sampling network design, Doctoral thesis, Department of mathematics and statistics, University of Vermont, Burlington, Vermont, 31 March.

Klir, G. J., Z. Wang, and D. Harmanec (1997), Constructing fuzzy measures in expert systems, *Fuzzy Sets and Systems, 92*, 251-264.

Michalak, A. M., and P. K. Kitanidis (2004), Estimation of historical groundwater contaminant distribution using the adjoint state method applied to geostatistical inverse modeling, *Water Resour. Res., 40*, W08302, doi:10.1029/ 2004WR003214.

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 (2001), Adjoint-derived location and travel time probabilities for a multidimensional groundwater system, *Water Resour. Res., 37*(6), 1657-1668.
Neupauer, R. M. (2004), Identifying sources of sorbing solutes in groundwater using backward location probabilities conditioned on measured concentrations, *Proceedings Computational Methods in Water Resources*, Chapel Hill, USA, 1251-1262.
Skaggs, T. H., and Z. J. Kabala (1994), Recovering the release history of a groundwater contaminant, *Water Resour. Res., 30*(1), 71-79.
Wagner, B. J. (2002), Simultaneous estimation and contaminant source characterization for coupled groundwater flow and contaminant transport modeling, *J. Hydrol., 35*, 275-303.
Zhang, Y., and G. F. Pinder (2000), A latin-hypercube method for evaluation of hydraulic conductivity random fields, *Proceedings Computational Methods in Water Resources,* 2, 779-784.