

Simulation of the Fate of Contaminant in Groundwater under Uncertainty Using Optimized Linear Interpolation

Z. Y. Hu*, C. W. Chan and G. H. Huang

Faculty of Engineering, University of Regina, Regina, Saskatchewan S4S 0A2, Canada

ABSTRACT. Subsurface characterization is an important requirement in developing a simulation model of the processes of subsurface contaminant transport and degradation. This paper presents a methodology that can model the inherent uncertainty that arises due to simplification of the processes and assumptions made about site data. The linear interpolation method was adopted to represent the changes in the hydraulic characteristics under study at uncertain locations within the subsurface site. An optimization method was designed to identify the proper range of those uncertain locations within which the hydraulic characteristics were assumed to change. A 3D contaminant fates and transport simulation model – UTCHEM (University of Texas at Austin, 2000) was used in this study to predict the site status. The solution method for this nonlinear optimization function was the multi-level cubic spline function. Results from the multi-level cubic spline function were compared with results generated from the fitting polynomials in the 6th and 9th orders. This method was used to analyze some data from a laboratory experiment. The results indicated that the optimized linear interpolation function can model this kind of uncertainty and enhance accuracy of the simulation model.

Keywords: Cubic spline function, optimized linear interpolation, simulation

1. Introduction

The first step in the design of a remediation strategy for a petroleum contaminated site is to use a multiphase and multicomponent model to simulate the passive and enhanced bioremediation processes in the subsurface. In this step, accuracy of the simulation model directly affects quality of the decision making. However, uncertainty is inherent and unavoidable in some of the input parameters in the simulation and accuracy of the simulation model can be compromised because of this uncertainty.

Previously, Larue (1997) suggested that the uncertainty inherent in hydrogeological modeling could originate from both randomness and vagueness of the parametric values. To handle the uncertainty that is caused by the vagueness of the parametric values, fuzzy set theory, probability analysis and stochastic analysis have been adopted. Different approaches have been adopted for modeling uncertainties in the problem. For example, Christakos et al. (1998), Zhu and Sykes (2000) and Freeze et al. (1990) used the stochastic method for evaluating uncertainties in groundwater flow and transport; Dou et al. (1997), Schulz and Huwe (1999), Schulz et al. (1999) and Zou and Lung (2000) applied fuzzy-set theory for modeling the imprecise parameters involved in groundwater transport; Abdin et al. (1996), Davis and Keller (1997) Naff, (1998a, b) and Hu et al. (2002) solved the simulation model using numerical methods based on the probability distributions of the input parameters. Li (2003) made a detailed re-

view of the different approaches for modeling uncertainties in the problem. But scant attention has been focused on modeling the uncertainty that arises due to randomness of parametric values. In this paper, the Optimized Linear Interpolation Method will be developed to model the uncertainty derived from randomness of parametric values. In the following, the source of uncertainty, the methodology for modeling this kind of uncertainty and the solution method will be presented in detail, and then a case study and some result analysis will be discussed.

2. Sources of Uncertainty

Subsurface characterization is an important requirement in simulating the processes of subsurface contaminant transport and degradation. For a contaminated site, heterogeneity appears in both site hydraulic properties and contaminant transport processes. However, the heterogeneity of aquifer properties is almost always poorly understood. Current field investigation methods do not adequately define the aquifer properties.

In remediation a real-world petroleum contaminated site, often remediation decisions are made based on a random, systematic or stratified sample because the site is typically so large that it is impossible to collect data on the entire area. Due to inadequacy of the sampled data and uncertainty in the assumptions of aquifer properties, heterogeneity at the site cannot be adequately and deterministically characterized. As a consequence, simulation of the contaminant transport and degradation is associated with inherent uncertainties. Hence, it

* Corresponding author: hzhyl1@uregina.ca

is important to develop a methodology that can determine the hydraulic characteristics such as soil particle size distribution, hydraulic conductivity, and contaminant concentration distribution of the entire site from limited soil sampling data. The Optimized Linear Interpolation Method has been developed for this purpose.

3. Optimized Linear Interpolation Method

3.1. Linear Interpolation

To clearly illustrate the method, distribution of the soil particle size is being analyzed here as an example. For all of the other site characteristics, the same method can be applied.

For a real-world contaminated site, the type of subsurface soil at different locations is likely to be different; however, between any two locations, the change in the soil particle size distribution can in fact be continuous. That is, the change in soil particle size distribution between two adjacent sampling points that have different soil types might be gradual. Therefore, it is not reasonable to represent the transition zone between different soil types at two sampling points as a crisp boundary between two different soil type zones. Since only isolated samples are taken from a site, describing the soil distribution and other hydraulic characteristics of the entire site requires making assumptions for areas from which no data have been collected. The assumption made is that the degradation occurs between soil type changes. As well, the hydraulic characteristics which are solely affected by media properties are also assumed to change gradually between any two sampling points.

The Optimized Linear Interpolation Method has been used to describe distribution of the soil particle size for locations from which no sampling data were collected. In a 2-D figure, $A(x_0, y_0)$, $B(x_2, y_0)$, $C(x_2, y_2)$ and $D(x_0, y_2)$ are the sampling points; while $M(x, y)$ is an unknown point located in the degradation belt among these four sampling points (Figure 1). The degradation between any two contiguous sampling points is assumed to be linear, which is a reasonable assumption because soil type degradation is normally linear in the real world subsurface. When the required accuracy of soil type classification is not very high, error due to this assumption is acceptable. The volume percentage of sand (R) at point $M(x, y)$ can be determined by using linear interpolation as follows:

$$R = R_a \times \frac{(y - y_0)}{(y_2 - y_0)} + R_b \times \frac{(y_2 - y)}{(y_2 - y_0)} \quad (1)$$

where

$$R_a = R_b \times \frac{(x - x_0)}{(x_2 - x_0)} + R_c \times \frac{(x_2 - x)}{(x_2 - x_0)} \quad (2)$$

$$R_b = R_c \times \frac{(x - x_0)}{(x_2 - x_0)} + R_d \times \frac{(x_2 - x)}{(x_2 - x_0)}$$

The same methodology can then be used for calculating

the volume percentages of clay and silt, and any other media properties under study.

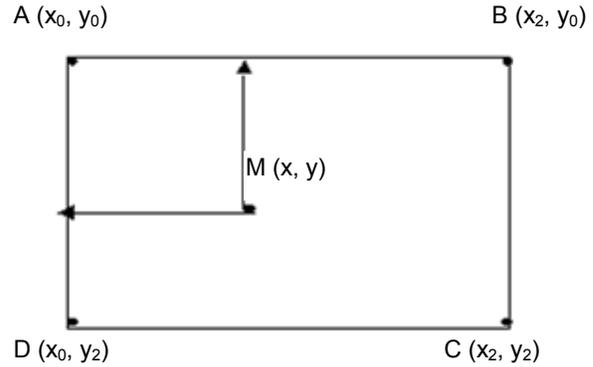


Figure 1. Sketch map of the location of samples and unknown point.

This method assumes that degradation regions exist between any two contiguous sampling points. In these degradation regions, the soil distribution and hydraulic characteristics solely affected by the media properties will degrade linearly from one sampling point to its neighbour. If a proper width of the degradation belt or region can be determined, the accuracy of the modeling results will be improved. The optimization method proposed in this study is used for determining the proper width of the degradation belt.

3.2. Optimization for the Width of Degradation Belt

When using the Linear Interpolation Method to calculate the value of an unknown point which is surrounded by some available sampling points, it is likely that the soil type around those sampling points are similar to the sampling points, but the soil type at points which are far away from those sampling points are less likely to be similar to those points. The proper width of the degradation belt is mainly affected by characteristics of the site itself. To more accurately simulate subsurface conditions at the concerned site, an optimization function is used. The Optimized Linear Interpolation Method is described as follows.

First, we can define the width of the degradation belt (w_1). Assuming a 1-D domain (x direction only), and $A(x_0, y_0)$, $B(x_2, y_0)$ are two sampling points, the distance between these two sampling points is: $AB = |x_0 - x_2|$.

Then for sampling point A, all points located along the line AB within the section $(x_0, x_0 + |x_0 - x_2| (1 - w_1)/2)$ have the same soil particle size as sampling point A, and all points located along the line AB with the section $(x_2 - |x_0 - x_2| (1 - w_1)/2, x_2)$ have the same soil particle size as sampling point B. For those points located within the sections $(|x_0 - x_2| (1 - w_1)/2 + x_0, y_0)$ to $(x_2 - (|x_0 - x_2| (1 - w_1)/2), y_0)$, the soil particle size changes gradually, as shown in Figure 2.

The soil particle size distributions in the entire site can be obtained using this method. As well, before inputting any

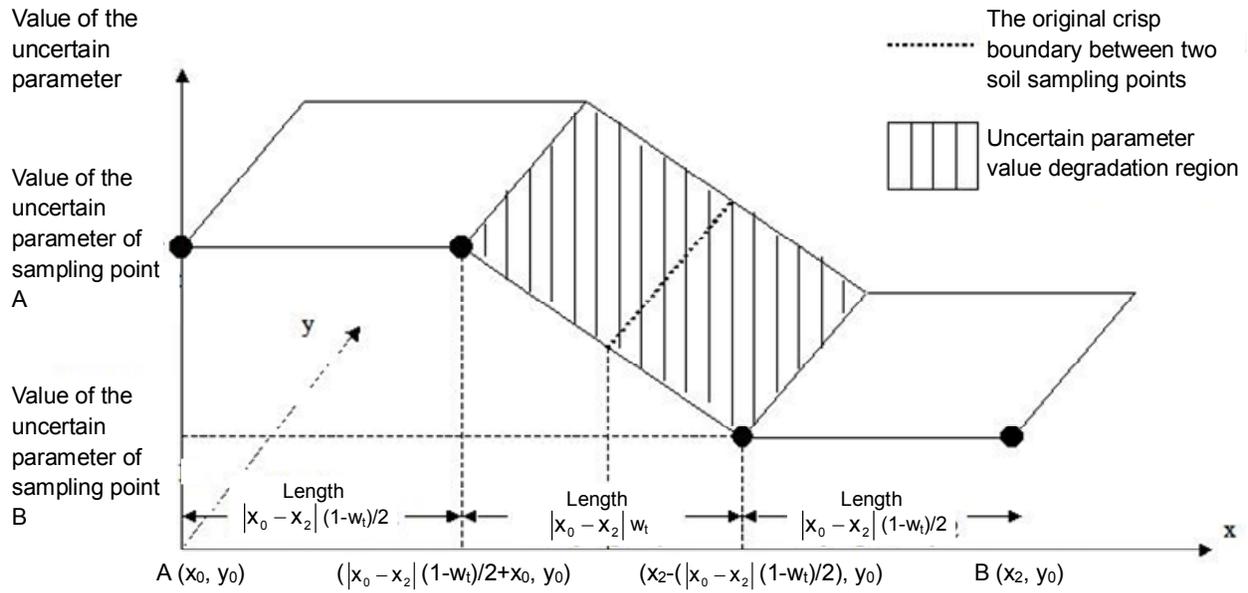


Figure 2. The width of the degradation belt (w_t).

other parameters under study into the simulation model, this method can also be used to handle this kind of uncertainty associated with them.

Since the value of w_t affects accuracy of the simulation model, the optimization function was used to identify the proper w_t value. After pre-treating all concerned uncertain parameters by the linear interpolation degradation belt method, we selected the proper value of the width of the degradation belt (w_t), in order to improve accuracy of the simulation result.

Let us assume all other input parameters are certain, so that the simulation result is affected by the width of the degradation belt (w_t) only. The objective function aims to minimize the difference between the simulation results and the monitored real site data, here the simulation results are the calculated contaminant concentration at each monitored point. The least square method has been used as the objective function, which serves to minimize the overall relative error between the calculated value and the monitored value. The objective function is generated as follows:

Objective:

$$J = \min \sum_{d=1}^M \left(\sum_{i=1}^N \left(\frac{C_{i,t+d} - C'_{i,t+d}}{C_{i,t+d}} \right)^2 \right) \quad (3)$$

Subject to:

$$0 < w_t \leq 1 \quad (4)$$

where $C_{i,t+d}$ is the measured contaminant concentration at

point i at time $t + d$; $C'_{i,t+d}$ is the calculated concentration value at point i at time $t + d$; the simulation model UTCHEM {University of Texas at Austin, 2000} has been used to calculate $C'_{i,t+d}$; i is the monitored point at the concerned site, in total, there are N monitored points; d is the monitored time, in total, there are M times the monitored data will be used in the function; M is decided by the user and affected by data availability.

4. Solution through Multi-Level Spline Functions

Objective (3) is a constrained nonlinear function. Since the relationship between w_t with $C'_{i,t+d}$ is nonlinear and complex, it is difficult to solve this function using analytic calculation methods. In this study, a numerical calculation method was used to generate the optimal w_t value.

In numerical calculation with a digital computer, it is necessary to consider problems of interpolation between given data points. One of these interpolating functions is known as a spline function. Schönberg first introduced the concept of Spline function in 1946 (Schönberg, 1946). It was used in many applications such as interpolation, data fitting, numerical solution of ordinary and partial differential equations, and in curve and surface fitting. Spline functions have the useful properties of being (1) smooth and flexible, (2) easy to store and manipulate on a computer, (3) easy to evaluate, along with their derivatives and integrals, and (4) easy to generalize to higher dimensions.

A multi-level spline function method has been adopted in this study, because it is difficult to select the number of knots that represent the curvature believed to correctly represent the data. In order to avoid generating only a local optimal result,

minated by petroleum hydrocarbons; the well locations and soil type distribution of the experimental tank is shown in Figure 4. The monitored concentrations of benzene from day 2 to day 10 are listed in Table 1. The measured and estimated intrinsic permeability (k) and porosity values of each grid are listed in Table 2.1 and Table 2.2. Since only some of the grids in this Table were sampled and measured in the experiment, the concerned hydraulic characteristics of the unknown grids were estimated based on the available sampling data.

Table 1. The Monitored Concentrations of Benzene from Day 2 to Day 10

| | Day2 | Day4 | Day6 | Day8 | Day10 |
|---------|-------|-------|-------|-------|-------|
| well 1 | 0.304 | 0 | 0 | 0 | 0 |
| well 2 | | | | | |
| well 3 | 1.926 | 1.593 | 1.429 | 0.508 | 0.401 |
| well 4 | 0 | 0 | 0.090 | 0.400 | 0.285 |
| well 5 | 1.497 | 1.920 | 2.070 | 1.985 | 0.824 |
| well 6 | 0.444 | | | 0.472 | 0.265 |
| well 7 | 0.385 | 0.733 | 1.227 | 0.686 | 0.300 |
| well 8 | 0.524 | 0.703 | 0.366 | 0.527 | 0.359 |
| well 9 | 1.270 | 1.202 | 1.398 | 1.257 | 1.083 |
| well 10 | 1.366 | 1.628 | 0.947 | 1.590 | 2.055 |
| well 11 | 0.649 | 0.710 | 0.663 | 0.554 | 0.463 |
| well 12 | 0.417 | 0.741 | 0.166 | 0.443 | 0.322 |

For a detailed explanation on the input file of the other site data (Huang, 2004). As shown in Figure 4, the boundaries between different soil types are crisp in each layer, which

means that the particle size combinations of those grids alone and at different side of each boundary changed suddenly. But in the real world, the soil particles distributions at the subsurface are continuous; so it is reasonable to assume that a degradation belt exists, in which the soil types changed gradually. In the simulation model used in this study, the input parameters of intrinsic permeability and porosity are functions of the media properties only. When linear interpolation is used to represent the degradation of the soil particle size distribution in one layer between two soil types, the degradation of porosity can be assumed linear also, and the intrinsic permeability is a function of the pore diameter, with the unit of L^2 , defined as:

$$K = Cd^2 \tag{23}$$

where K is intrinsic permeability (L^2); C is shape factor properties of Media; and d is mean pore diameter (L).

Under the assumption that within the degradation belt, the volumes of the soil particles changed linearly, so the degradation of $k^{3/2}$ can be assumed linear also. To define the maximum width of the degradation region, the sampling data have been reviewed for this purpose and the width of the soil type degradation region has been defined as the distance between two sampling points which have different soil types along the X direction. For those regions that do not have detailed soil sampling data near the boundary between different soil types, the degradation belt is simply assumed to be equal to 4 grids (1.2 m) in the X direction, corresponding to w_x equal to 1.

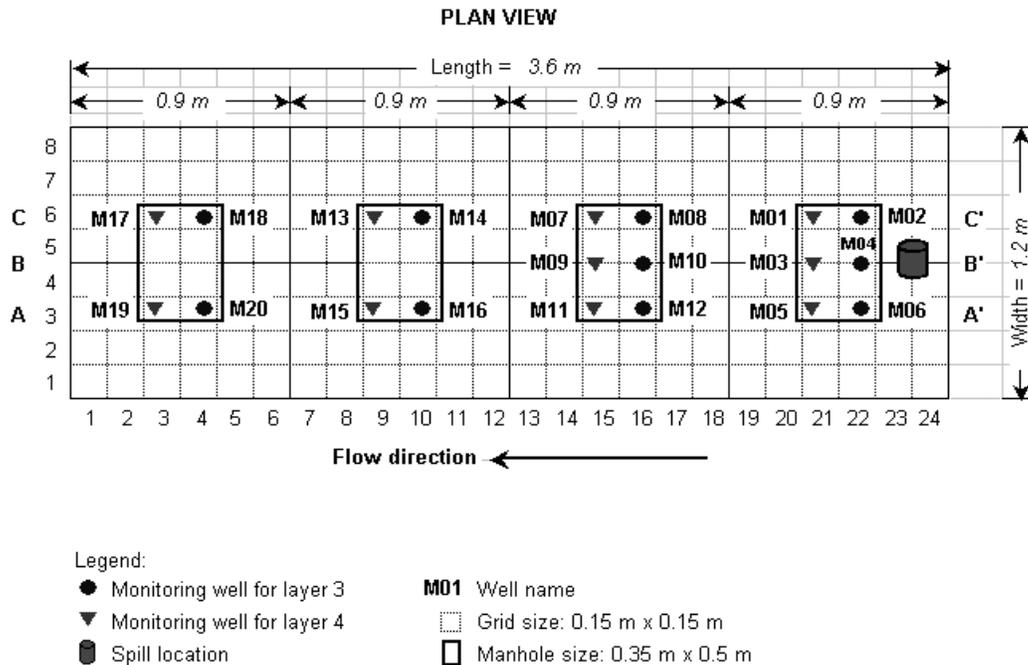


Figure 3. Tank plan view.

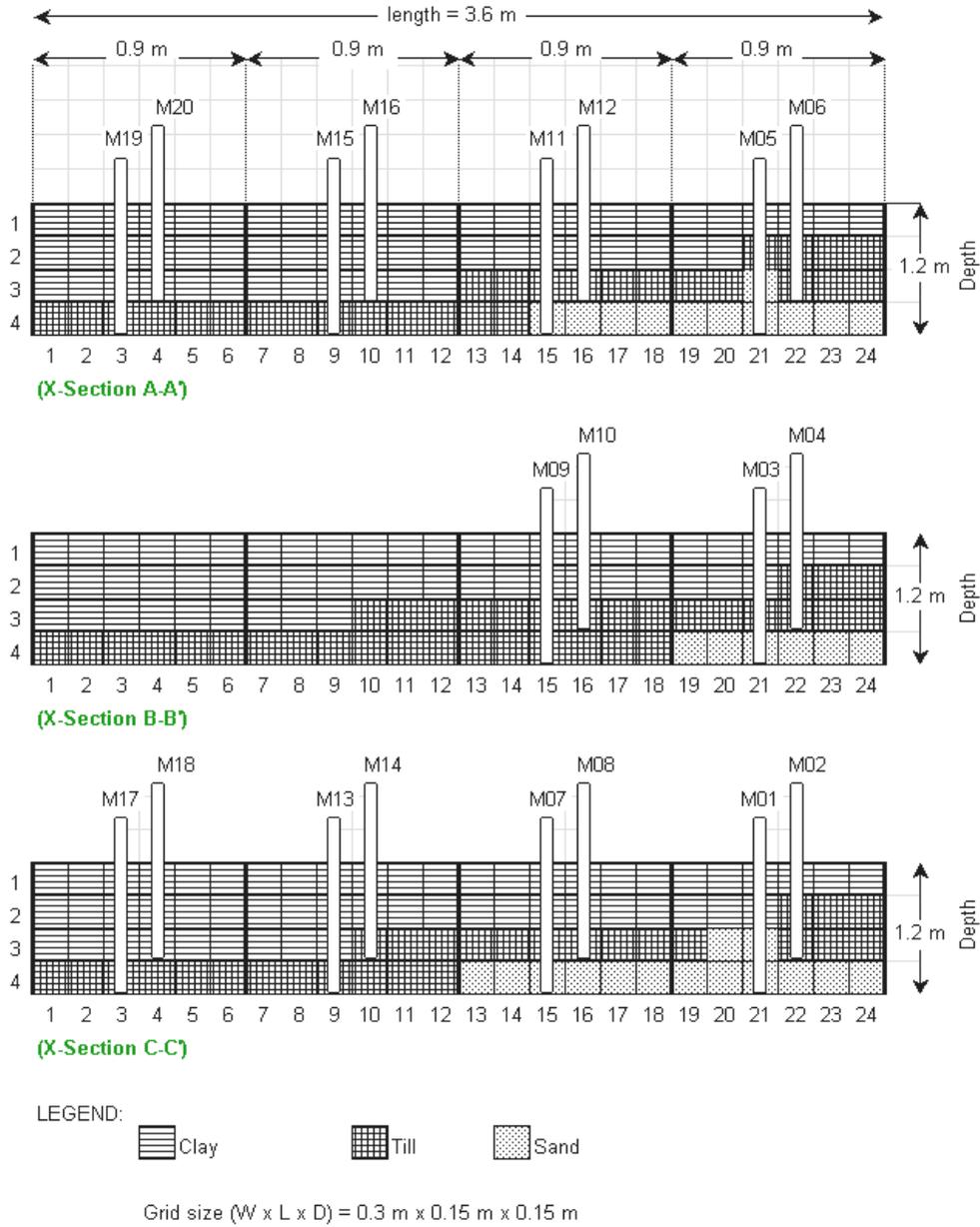


Figure 4. Well locations and soil types.

Table 3. Knots Values (w_t) and the Corresponding $J(w_t)$ Levels at Interested Wells during Days 2 to 6

| | | | | | | | | | | | |
|----------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| w_t | 0 | 0.05 | 0.1 | 0.15 | 0.2 | 0.25 | 0.3 | 0.35 | 0.4 | 0.45 | 0.5 |
| $J(w_t)$ | 2.2587 | 2.2798 | 2.2764 | 2.273 | 2.2699 | 2.2706 | 2.2683 | 2.2659 | 2.2659 | 2.2613 | 2.2588 |
| w_t | 0.55 | 0.6 | 0.65 | 0.7 | 0.75 | 0.8 | 0.85 | 0.9 | 0.95 | 1 | — |
| $J(w_t)$ | 2.2568 | 2.2547 | 2.2525 | 2.2505 | 2.2484 | 2.2464 | 2.2444 | 2.2428 | 2.2412 | 2.2853 | — |

Two sets of w_t values are selected first, $W_t(1) = (0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0)$, and $W_t(2) = (0, 0.05, 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85, 0.95, 1)$. Table 3 lists all w_t values and the corresponding $J(w_t)$ levels.

For set a, the calculated optimal w_t value $w_t(a) = 0.87$, which belongs to interval (0.8, 0.9), the corresponding $J(w_t)$ value is 2.2419. For set b, the optimal w_t value $w_t(b) = 0.92$, which belongs to interval (0.85, 0.95), the corresponding $J(w_t)$ value is 2.2348. Figure 5 shows the curves of sets a and b. In this figure, the dotted line represents the data for set b and the line with "x" represents those for set a.

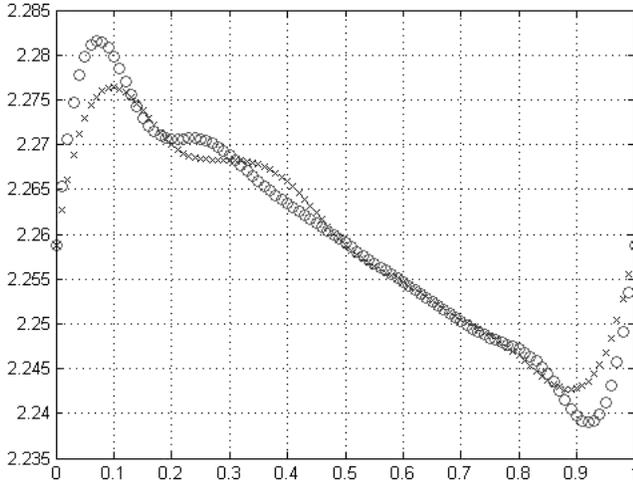


Figure 5. The fitting curves generated by cubic spline function of sets a and b.

Therefore new knots have been generated within region (0.8, 0.95) as $w_t = (0.82, 0.84, 0.86, 0.88, 0.91, 0.93)$. By using the multi-level cubic spline function described above, the optimal w_t value is 0.92 and the corresponding J level is 2.2372.

6. Result Analysis

The calculated optimal w_t value has been used to modify the input permeability and porosity values. The square of the relative errors of w_t equal to 0.92 ($\epsilon(0.92)^2$) at some monitoring wells has been calculated and compared with the corresponding values under the condition that no degradation belt was considered ($w_t = 0, \epsilon(0)^2$). Table 4 lists the difference between these values. It is shown that by using w_t to model the uncertainty, accuracy of the simulation result of the entire site ($\sum(\epsilon(0.92)^2 - \epsilon(0)^2)$) and most of the concerned monitoring wells can be improved. It can be observed that at some monitoring wells, the accuracy of the simulation result when

w_t is equal to 0.92 is worse than that for when w_t is equal to 0. This observation is reasonable and can be explained as follows: in a real-world site, since the distribution of soil characteristics might be different spatially, so different regions might have different proper w_t values. In this study, the optimal w_t value was calculated based on the assumption that the entire site has the same w_t value, so the optimal w_t value can only improve accuracy of the result of the entire site and not that for each monitoring well.

Table 4. Difference of the Squares of the Relative Errors under $w_t = 0.92$ and $w_t = 0$ for the Calculated Benzene Concentrations at Wells 8, 9, 10, 11 and 12 ($\epsilon(0.92)^2 - \epsilon(0)^2$)

| | Day 2 | Day 4 | Day 6 | Day 8 | Day 10 |
|---------|----------|----------|----------|----------|----------|
| well 8 | -0.01749 | -0.01229 | -0.05344 | -0.00051 | 0.004647 |
| well 9 | -0.00342 | -0.00392 | -0.00126 | -0.00221 | -0.00448 |
| well 10 | -0.00135 | -0.0015 | -0.0125 | -0.00039 | 0.000778 |
| well 11 | -0.00014 | -0.00063 | -0.00042 | 0.000667 | 0.001502 |
| well 12 | -0.05392 | -0.02912 | -1.05515 | -0.09236 | -0.1053 |

The result has also been compared with the results generated by the Monte Carlo method and fitting polynomial. The Monte Carlo method can generate the same result but required more than 100 times of executing the simulation model. By using the multi-level cubic spline function method presented in this paper, only 24 times of executing the simulation model was needed. Therefore, it has been demonstrated that using the cubic spline method can reduce substantial computer time. The 9th order fitting polynomial was also tested; the result is void because of the appearance of the Runge Phenomenon.

A sensitivity analysis was conducted and the values of $J(0.96)$ and $J(0.88)$ were calculated which correspond to w_t with 5% error, $0.92(1 \pm 5\%)$. As listed in Table 5, $J(0.96) = 2.2497$ and $J(0.88) = 2.2436$; the corresponding absolute relative errors of $ABS(J(w_t \pm error) - J(w_t))/J(w_t)$ are 6 and 0.3%, respectively. Thus it can be seen that the change of w_t value does affect accuracy of the simulation result. When the w_t value approximates but is less than the optimal w_t , then the effect of the error of w_t will be less than that of the w_t value which approximates but is larger than the optimal w_t .

7. Conclusions

In this paper, the Optimized Linear Interpolation Method has been developed to handle uncertainties that arise due to simplification and assumption of site data in the simulation of

Table 5. Sensitivity Analysis

| $J(0.92)$ | $J(0.92 - 5\% \text{ error})$ | $J(0.92 + 5\% \text{ error})$ | $(J(w_t - error) - J(w_t))/J(w_t)$ | $(J(w_t + error) - J(w_t))/J(w_t)$ |
|-----------|-------------------------------|-------------------------------|------------------------------------|------------------------------------|
| 2.2372 | 2.2436 | 2.2497 | 0.3% | 6% |

NAPL fate and transport in groundwater. The multi-level cubic spline function was used to solve the constrained non-linear optimization function. This method was applied on data from a tank at a laboratory. The result of the case study indicated that the method presented in this paper can handle the problem of uncertain data and enhance accuracy of the simulation results. This method is potentially applicable to other fields that involve this kind of uncertain data. An objective in future work is to enhance the methodology for the purpose of generating different w_i values at different locations, in an effort to improve accuracy of the simulation results for both the entire site and for each monitoring well.

Acknowledgments. We are grateful for the generous support of a Strategic Grant from the Natural Sciences and Engineering Research Council of Canada.

References

- Abdin, A., Kaluarachchi, J., Kembrowski, M.W. and Chang, C.M. (1996). Stochastic analysis of multiphase flow in porous media: II. numerical simulations. *Stochastic Hydrol. Hydraul.*, 10(3), 231-251.
- Ahilbero, J.H., Nilson, E.N. and Walsh, S.L. (1967). *The Theory of Splines and Their Applications*, Academic Press, New York.
- Bartels, R.H., Beatty, J.C. and Barsky, B.A. (1998). Hermite and cubic spline interpolation, in *An Introduction to Splines for Use in Computer Graphics and Geometric Modeling*, Morgan Kaufmann, San Francisco, CA, pp. 9-17.
- Christakos, G., Hristopulos, D.T. and Li, X.Y. (1998). Multiphase flow in heterogeneous porous media from a stochastic differential geometry viewpoint. *Water Resour. Res.*, 34(1), 93-102.
- Davis, T.J. and Keller, C.P. (1997). Modeling uncertainty in natural resource analysis using fuzzy sets and Monte Carlo simulation: slope stability prediction. *Int. J. Geogr. Inf. Sci.*, 11(5), 409-434.
- Dou, C., Woldt, W., Bogardi, I. and Dahab, M. (1997). Numerical solute transport simulation using fuzzy sets approach. *J. Contam. Hydrol.*, 27(1-2), 107-126.
- Freeze, R.A., Massmann, J., Smith, L., Sperling, J. and James, B. (1990). Hydrogeological decision analysis, 1, a framework. *Ground Water*, 28(5), 738-766.
- Hartree, D.R. (1958). *Numerical Analysis* (2nd Edition), Oxford University Press, London.
- Spline Function. <http://rkb.home.cern.ch/rkb/AN16pp/node272.html>.
- Hu, B.X., Huang, H. and Zhang, D. (2002). Stochastic analysis of solute transport in heterogeneous, dual-permeability media. *Water Resour. Res.*, 38(9), 14-1 - 14-16.
- de Boor, C. (1978). *A Practical Guide to Splines*, Springer Verlag, New York.
- Huang, Y.F. (2004). *Development of Environmental Modeling Methodologies for Supporting System Simulation, Optimization and Process Control in Petroleum Waste Management*, Ph.D. Dissertation, University of Regina, Regina, Canada.
- Larue, J.L. (1997). *A Fuzzy Modeling Platform for Flow and Transport in Alluvial and Fractured Porous Media*, Ph.D. Dissertation, Oklahoma State University, OK, USA.
- Li, J.B. (2003). *Development of an Inexact Environmental Modeling System for the Management of Petroleum Contaminated Sites*, Ph.D. Dissertation, University of Regina, Regina, Canada.
- Naff, R.L., Haley, D.F. and Sudicky, E.A. (1998a). High-resolution Monte Carlo simulation of flow and conservative transport in heterogeneous porous media, 1. methodology and flow results. *Water Resour. Res.*, 34(4), 663-677.
- Naff, R.L., Haley, D.F. and Sudicky, E.A. (1998b). High-resolution Monte Carlo simulation of flow and conservative transport in heterogeneous porous media, 2. transport results. *Water Resour. Res.*, 34(4), 679-697.
- Schoenberg, I.J. (1946). Contribution to the problem of approximation of equidistant data by analytic functions. *Quart. Appl. Math.*, 4, 45-99, 112-141.
- Schulz, K. and Huwe, B. (1999). Uncertainty and sensitivity analysis of water transport modelling in a layered soil profile using fuzzy set theory. *J. Hydroinf.*, 1(2), 127-138.
- Schulz, K., Huwe, B. and Peiffer, S. (1999). Parameter uncertainty in chemical equilibrium calculations using fuzzy set theory. *J. Hydrolog.*, 217(1), 119-134.
- Traur, J.F. (1964). *Iterative Methods for the Solution of Equations*, Prentice-Hall, Englewood Cliffs, NJ, USA.
- University of Texas at Austin (2000). Technical documentation for UTCHEM-9.0, *A Three-Dimensional Chemical Flood Simulator, Reservoir Engineering Research Program*, Center for Petroleum and Geosystems Engineering, University of Texas at Austin, TX, July 2000.
- Zhou, D.G., Dillard, L.A. and Blunt, M.J. (2000). A physically based model of dissolution of nonaqueous phase liquids in the saturated zone. *Transp. Porous Media*, 39(2), 227-255.
- Zhu, J. and Sykes, J.F. (2000). Stochastic simulations of NAPL mass transport in variably saturated heterogeneous porous media. *Transp. Porous Media*, 39(3), 289-314.