Uncertainty in spatial distribution of residual NAPL and its downstream impacts

Amir H Hosseini, Clayton V Deutsch
Department of Civil and Environmental Engineering, University of Alberta, Edmonton, AB, Canada
Kevin W Biggar
BGC Engineering Ltd., Edmonton, AB, Canada
Carl A Mendoza
Department of Earth and Atmospheric Sciences, University of Alberta, Edmonton, AB, Canada

ABSTRACT
In spite of several sophisticated models proposed to quantify the rate of mass transfer between residual non-aqueous phase liquid (NAPL) and groundwater, little attention has been paid to the geometry of residual NAPL source zone in a fate and transport modeling context. Uncertainty in spatial distribution of residual NAPL often affects the size and longevity of predicted dissolved contaminant plume in groundwater. In this work, first, a distance-function based approach is presented to quantify the uncertainty in the areal limits of a residual NAPL plume at a site impacted by petroleum hydrocarbons. In the second step, multiple secondary data sources such as soil texture and groundwater table elevation are combined and used in generating multiple geostatistical realizations of the presence and absence of contaminants. The 3D geostatistical realizations are then clipped by 2D realizations of areal extent to give the final model of uncertainty for residual NAPL distribution. A cross validation exercise is implemented to show the value of the secondary data in improving the prediction ability in the proposed methodology. Finally, results from a set of Monte Carlo Simulations shows the impact of uncertain source zone geometry on downstream concentrations for a moderately heterogeneous aquifer.

RÉSUMÉ
En dépit de plusieurs modèles sophistiqués proposé de quantifier le taux de transfert de masse entre résiduels non liquides phase aqueuse (NLPA) et les nappes phréatiques, peu d'attention a été accordée à la géométrie de la NLPA résiduel dans une zone de la source du devenir et du transport de modélisation contexte. Incertitude dans la répartition spatiale de la NLPA résiduel souvent influe sur la taille et la longévité des prédit panache de contaminants dissous dans les eaux souterraines. Dans ce travail, d'abord, sur une distance de l'approche fondée sur la fonction est présenté pour quantifier l'incertitude dans les limites d'une superficie résiduelle NLPA panache à un endroit touché par les hydrocarbures pétroliers. Dans la deuxième étape, de multiples sources de données secondaires telles que la texture du sol et l'élévation de la nappe phréatique sont combinés et utilisés en générant de multiples réalisations géostatistiques de présence et l'absence de contaminants. Les réalisations géostatistiques 3D sont ensuite clipser par 2D étendue des réalisations de donner le modèle final de l'incertitude pour la distribution de NLPA. Une croix exercice de validation est mis en œuvre pour démontrer la valeur des données secondaires pour améliorer la capacité de prévision dans la méthode proposée. Enfin, les résultats d'une série de simulations de Monte Carlo montre l'impact de l'incertitude sur la géométrie de la zone source du panache dissous taille et la forme d'une pour classe hétérogène aquifère.

1 INTRODUCTION

Light non-aqueous phase liquids (LNAPLs) associated with petroleum hydrocarbons, are typically produced, stored and distributed as gasoline, diesel, heavy fuel, and lubricating oil. Many groundwater contamination incidents begin with the release of these immiscible fluids into the vadose zone. The pattern of movement, distribution and redistribution of mobile and residual LNAPL in the subsurface is closely related to the soil texture and the dominant hydrogeological regime at the site. When LNAPL is released, it migrates vertically and laterally under the gravity and capillary forces. The LNAPL migrates through the unsaturated zone toward capillary fringe and the water table (Figure 1). Due to capillary forces, some LNAPL is always retained in the soil pores as 'residual' NAPL. As the remaining 'mobile' LNAPL continues to migrate through the subsurface, the volume of the mobile product decreases and LNAPL becomes trapped as isolated droplets within the soil pore network. In other words, NAPL plumes are spatially 'self-limiting', unless continually supplied from an ongoing release. While migrating through the subsurface, LNAPL distribution is affected by the heterogeneous nature of the soil. Slight differences in soil texture may promote preferential pathways horizontally and vertically. Residual LNAPL distribution and redistribution is also significantly influenced by seasonal fluctuations of groundwater table. Removal of residual NAPL is extremely difficult. It slowly dissolves into groundwater and creates a long-term source of groundwater pollution.

Partitioning of NAPL into aqueous and vapour phases and dissolution of non-uniformly distributed NAPL into groundwater have been extensively studied in the last two

decades (Imhoff et al. 1993, Brusseau et al. 2002). Nevertheless, there is no systematic approach to evaluate the geometry and structure of residual NAPL source and its uncertainty in a field-scale.

In this paper, a two-step geostatistical approach is proposed to delineate the space of uncertainty for distribution of residual NAPL in a contaminated aquifer. First, a distance-function based approach is used to characterize the uncertainty in areal limits of the contaminant source zone. In the second step, multiple secondary data sources are combined and used in generating 3D realizations of presence/absence of residual NAPL. These realizations are eventually clipped by 2D realizations of areal extent obtained in the previous step. The value of different secondary data sources in improving the prediction ability is then investigated by cross-validation. Results of a set of Monte Carlo Simulations are then reported to show the importance of modeling uncertainty in source geometry in a fate and transport modeling context. The proposed methodology is presented in the form of a case study for a hydrocarbon impacted site located in west-central Alberta, Canada.

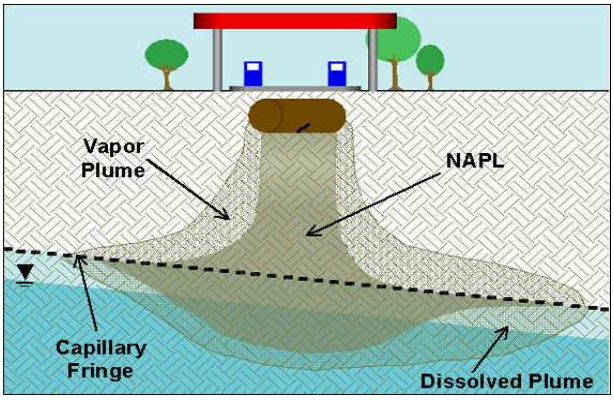

Figure 1. A schematic representation of LNAPL release in the subsurface.

2 DISTANCE FUNCTION APPROACH

Geostatistical estimation and simulation techniques (Deutsch and Journel 1998) are applied with Stationary Random Functions (SRF) within geometric limits imposed by areal (volumetric) limits. There is often significant global uncertainty in areal boundaries depending on the amount of data available. Point measurements of soil NAPL concentrations should not be directly used in geostatistical modeling of contaminant source zone unless the areal boundary of the stationary domain is reasonably defined and its uncertainty is characterized.

The definition of distance function (DF) is closely related to the notion of distance to an interface separating two distinct domains within which two different SRFs will subsequently be developed for geostatistical modeling. Distance is measured to the nearest interface. An initial binary coding of the available sample data in terms of being 'inside' and 'outside' the source zone is a prerequisite for constructing a DF. The distance function must be calculated for all data locations and control points for subsequent interpolation. Figure 2 shows how samples are coded for monitoring wells deemed inside and outside the source zone and how the DF is calculated as the distance to the nearest unlike data location. An initial guess for the boundary may be obtained by tracing the line corresponding to DF equal to 0.0. Besides the known well locations, an arbitrary number of control points can be added at the locations which are known a-priori to be inside or outside of the contaminated area (Figure 2).

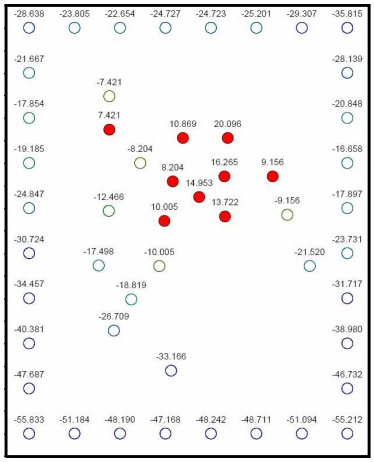

Figure 2. Configuration of contaminated (solid circles) and uncontaminated wells and calculation of DF values at all data locations and control points.

An interpolation technique is employed to define the boundary interface in presence of sparse sample data. As negative weights should be avoided in interpolation, an inverse distance approach is preferred to kriging. An inverse distance estimate at an unsampled location u_0 is a weighted linear combination of N surrounding sample data $Z(u_i)$ in a search neighbourhood and is given by the following expression:

$$Z^*_{ID}(u_0) = \sum_{i=1}^{N} \lambda_{ID}(u_i) \cdot Z(u_i) \qquad [1]$$

The weights $\lambda_{ID}(u_i)$ are calculated by:

$$\lambda_{ID}(u_i) = \frac{\frac{1}{(d(u_i))^{\omega} + c}}{\sum_{j=1}^{N} \frac{1}{(d(u_j))^{\omega} + c}} \qquad [2]$$

where, $d(u_i)$ is the Euclidian distance between the estimation location u_0 and $Z(u_i)$ sample data, ω is distance exponent and c is a constant. The exponent ω controls the smoothness of the inverse distance estimates. Often, it takes a value between 0.5 and 2.0. Similar to nugget effect in the variogram model used for kriging, the constant c controls the short-scale variability in the estimates, and should be set to very small value. The search neighbourhood for inverse distance interpolation is often calibrated by cross-validation to limit

the number of data used in interpolation (Rojas-Avellaneda and Silvan-Cardenas 2006). In the context of the DF algorithm, however, it is recommended to use large searches to ensure smooth estimates. DF conditioning data must be honoured exactly.

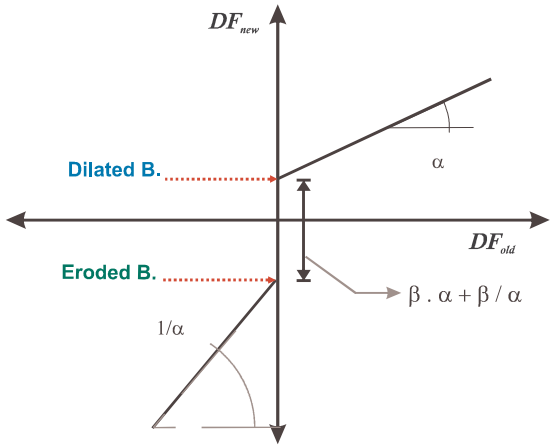

Figure 3. Conditioning the DF data with scaling and separation factors.

The DF is not a stationary random function and traditional probabilistic approaches cannot be used for uncertainty assessment. Kriging and inverse distance weights are typically dependent on the geometrical arrangement of the data locations, not the actual data values. In this work, a data value dependent weighting scheme (Figure 3) is presented by using a general parameterization scheme for data conditioning:

$$Z^{*}_{ID}(u_0) = \sum_{i=1}^{N} \alpha^{\frac{Z(u_i)}{|Z(u_i)|}} \cdot \lambda_{ID}(u_i) \cdot \left[Z(u_i) + \beta \cdot \frac{Z(u_i)}{|Z(u_i)|} \right] \qquad [3]$$

in which, α and β are scaling and separation factors, respectively. All other terms have been defined previously. The scaling factor α can take any value greater than zero. Increasing or decreasing α results in dilated large or eroded small boundary domains. The separation factor β is the other data conditioning factor that is closely related to the notion of uncertainty band and can take any value greater than or equal to zero.

For a given well arrangement, an 'uncertainty band' is defined as a probabilistic areal interval that includes the actual boundary which is unknown. The factors α and β control the band-width and the location of the centerline of the uncertainty band. They should be calibrated to appropriately represent the space of uncertainty for areal limits. The calibration is implemented using a large number of synthetic plumes. Every time, a number of wells are randomly added to the existing well configuration and the new setting is used to create a synthetic plume. First, all observation wells and control points are coded as either 1 (contaminated) or 0 (uncontaminated), and a directional search angle, θ, is specified. Next, for every contaminated well, a directional search is implemented and the directions that include a closest 'unlike' data location are identified as 'valid' search directions and those that include a closest 'like' data location are identified as null search directions (Figure 4). Then, for each contaminated well a search direction is randomly selected. If the selected search direction is valid, a new imaginary well is added to the setting. In essence, this imaginary well is randomly located on a line that connects the original contaminated well to the closest uncontaminated well location, and it is randomly coded as either contaminated or uncontaminated. The DFs are then recalculated for the new setting and mapped by inverse distance interpolation. The line of DF = 0.0 is then traced to create a new realization. The size of the search angle, its starting orientation and search neighbourhood for interpolation are important parameters in this process. A large number of synthetic realizations created using this approach fully represents the space of uncertainty associated with areal limits.

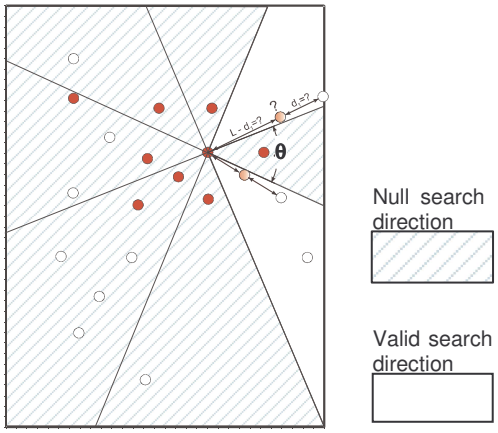

Figure 4. Implementation details of directional search in generating synthetic realizations.

As explained before, α and β control the centerline and the width of uncertainty band and should be calibrated simultaneously against the generated realizations. The centerline is calibrated to ensure unbiasedness and the width is calibrated to ensure fair probability distribution. This is a problem of optimization with the following objective function to be minimized:

$$S(\alpha,\beta,R) = \sum_{j=q_1}^{q_M} \left[P_j - P_j^{*}(\alpha,\beta,R) \right]^2 \qquad [4]$$

where, P_j represents the true probabilities corresponding to quantiles $q_1,...,q_M$ used in optimization. The calculated probabilities $P_j^{*}(\alpha,\beta,R)$ are defined as the proportion of synthetic plumes (R) that their areas completely fall inside the $q_1,...,q_M$ quantile maps. These quantile maps are derived from the conditional cumulative distribution function (CCDF) of an uncertainty band calculated for some α and β values. In this work, downhill simplex optimization algorithm has been adapted. The calibrated

values of α = 1.36 and β = 12.92 minimize the objective function presented in equation 4. The calibrated uncertainty band and corresponding quantile maps are presented in Figure 5.

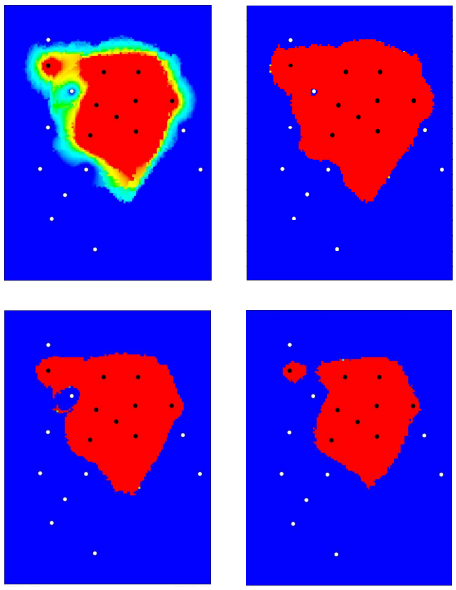

Figure 5. The calibrated uncertainty band (top-left) and p90 (top-right), p50 (bottom-left), and p10 (bottom-right) quantile maps.

3 SECONDARY DATA INTEGRATION

Data from multiple data sources such as soil texture and groundwater surface elevation can be combined with the assumption of conditional independence. This gives rise to a 3D map for conditional distribution of presence-absence of contamination conditioned to secondary data sources. In a sequential indicator simulation (SIS) context (Goovaerts 1997), this conditional distribution is combined with prior probability map to build a 3D updated posterior probability map. In this work, indicator hard data as well as soil texture data come from Ultra-Violet Induced Fluorescence Cone Penetration Testing (CPT-UVIF). Groundwater elevation data are obtained from 23 piezometers installed at the contaminated site.

3.1 Primary hard data

CPT-UVIF has been frequently used in environmental site characterization. Commercially available CPT-UVIF is a standard CPT cone coupled with the UVIF module to detect zone impacted by aromatic hydrocarbons. The cone records the mechanical responses of the soil at the same scale as recorded UVIF signals. The UVIF responses can be only reliably used as a screening tool to identify contamination by LNAPLs. In other words, the UVIF response is an indictor of presence/absence of contamination and can easily be incorporated into geostatistical modeling. In this work a categorical variable (T-UVIF) is introduced to represent the presence or absence of contamination based on the UVIF responses:

$$i(u_\alpha;k)=\begin{cases}1, & \text{if LNAPL is present } (k=1) \text{ at location } u_\alpha \\ 0, & \text{otherwise } (k=0)\end{cases} \quad [5]$$

The global proportions for presence or absence of contamination are 0.267 and 0.733, respectively.

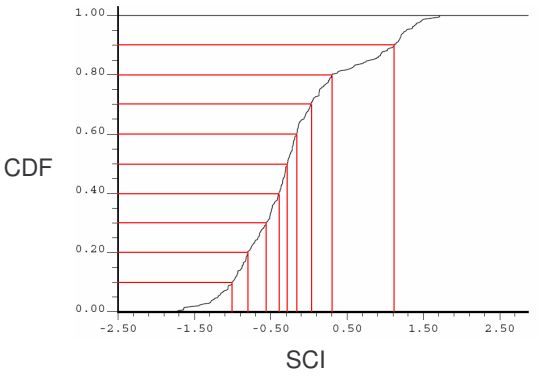

Figure 6. The CDF of the SCI data with 10 equal-sized classes defined by the deciles

3.2 Secondary soft data: soil texture

One of the most important aquifer properties affecting residual NAPL distribution is soil heterogeneities. Short-range variability in soil properties results in preferential flow. In a heterogeneous setting, the distribution of NAPL is generally correlated to the distribution of higher porosity and more permeable units. In this study, cone penetration testing (CPT) data have been used to model the geological structure at the site. The CPT instrument continuously records the mechanical response of the soil at a high resolution. Following the methodology introduced by Zhang and Tumay (2003), soil classification index (SCI) can be calculated at every data location. Figure 6 shows the cumulative distribution function (CDF) of SCI data. The presence/absence of contamination (from UVIF output) is calibrated against the SCI data and a calibration table is established (Table 1).

Table 1. Calibration of SCI based on the observed T-UVIF

SCI – class (y_{SCI})	$p(k=1\|y_{SCI})$	$p(k=0\|y_{SCI})$
[-2.14,-1.01]	0.093	0.907
[-1.01,-0.8)	0.1515	0.8485
[-0.8,-0.56)	0.2	0.8
[-0.56,-0.39)	0.2424	0.7576
[-0.39,-0.29)	0.2353	0.7647
[-0.29,-0.16)	0.3243	0.6757
[-0.16,0.04)	0.2632	0.7368
[0.04,0.31)	0.3429	0.6571
[0.31,1.11)	0.3902	0.6098
[1.11,1.77)	0.4211	0.5789

A positive correlation is observed in table 1 between the two attributes. To generate a 3D map of conditional probabilities, $p(k|y_{SCI})$, 100 realizations of the SCI field are generated by sequential Gaussian simulation (Deutsch and Journel 1997) on a 120 × 160 × 56 grid.

The equal-sized cell dimensions are 0.5m × 0.5m × 0.25m. Appropriate conditional probabilities (from Table 1) are assigned to each cell in every realization and then averaged over all realizations.

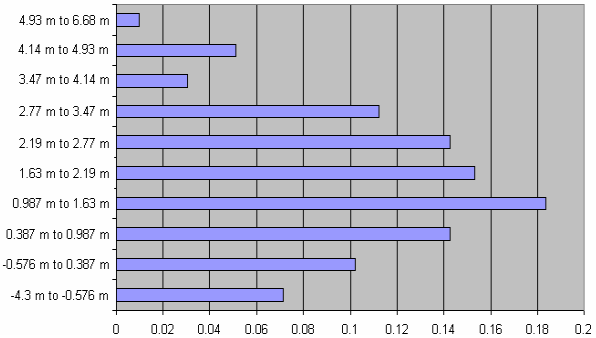

Figure 7. Global proportions of contamination for different classes of normalized elevation

3.3 Secondary soft data: location relative to GW table

The vertical movement of groundwater table affects the volume of mobile and residual LNAPL. Given some mobile LNAPL sitting on the groundwater table, a rise in groundwater table elevation causes the hydrocarbon to migrate upward as groundwater displaces it from the pore space. As water fills the pore network, LNAPL becomes trapped in the form of small droplets. These isolated droplets remain suspended in the network until the water table elevation drops. Lowering the water table enables the LNAPL drain from the pore network. During drainage, droplets of LNAPL remain within the pore interfaces leaving residual LNAPL within the unsaturated zone. The resultant vertical movement of the water table produces a residual 'smear zone' within the saturated and unsaturated zones.

Table 2. Calibration of Z_{normal} based on observed T-UVIF

Z_{normal} – class (y_{GW})	$p(k=1\|y_{GW})$	$p(k=0\|y_{GW})$
[-4.3m,-0.57m]	0.212	0.788
[-0.57m , 0.38m)	0.294	0.706
[0.38m ,0.987m)	0.326	0.674
[0.987m ,1.63m)	0.461	0.539
[1.63m ,2.19m)	0.384	0.616
[2.19m ,2.77m)	0.326	0.674
[2.77m ,3.47m)	0.333	0.667
[3.47m ,4.14m)	0.083	0.917
[4.14m ,4.93m)	0.151	0.849
[4.93m ,6.68m)	0.029	0.971

To account for the effects of groundwater table fluctuations, a parameter 'normalized elevation' is defined as the elevation of T-UVIF data point relative to groundwater table elevation at the same location:

$$Z_{normal} = Z_{UVIF} - Z_{GW} \quad [6]$$

The presence/absence of contamination is calibrated against Z_{normal} data and conditional probabilities are calculated. Figure 7 shows the global probabilities of contamination for different classes of normalized elevation. Calibration of absence/presence of contamination has been summarized in Table 2.

3.4 Integration of secondary data sources: permanence of ratios

A robust approach is to assume the data is conditionally independent given the primary data event. The expression for conditional probability of the primary data event k given the secondary data events y_{SCI} and y_{GW} is:

$$p(k|y_{SCI}, y_{GW}) = \frac{p(k) \times p(y_{SCI}|k) \times p(y_{GW}|k)}{p(y_{SCI}, y_{GW})} \quad [7]$$

where, the joint probability $p(y_{SCI}, y_{GW})$ is required. According to Journel (2002), Bayesian analysis goes around this problem by considering the ratios of the updated probabilities of the type. This results in the permanence of ratios (PR) assumption. Thus, the expression for the conditional probability based on the assumption of permanence of ratios (conditional independence) is expressed by:

$$p(k|y_{SCI}, y_{GW}) = \frac{\frac{p(\tilde{k})}{p(k)}}{\frac{p(\tilde{k})}{p(k)} + \frac{p(\tilde{k}|y_{SCI})}{p(k|y_{SCI})} \cdot \frac{p(\tilde{k}|y_{GW})}{p(k|y_{GW})}} \quad [8]$$

in which, the event $\tilde{k}$ represents the complement of the primary data event k. The secondary sources of information (y_{SCI}, y_{GW}) are combined using Equation 8 to create a 3D map for conditional probability distribution of presence/absence of contamination. This conditional probability distribution is then used in the next sub-step to find the updated posterior distribution of residual NAPL.

3.5 Integration of the prior probability distribution with the conditional probabilities

The conditional probabilities obtained in previous steps should be integrated to the prior probability map, which is conditioned to hard data only and built using the SIS. There are a number of techniques used to constrain the SIS to soft secondary data. In this work, two different techniques are used and their results are compared. First, we implement sequential indictor simulation with locally varying means (SIS-LVM). Bayesian Updating (BU) is performed next. Performance of the two approaches is then compared by cross-validation.

The conditional probabilities can be incorporated as the locally varying means for simulation. Therefore, the expression for probability of presence or absence of contamination can be written by (Deutsch 2006):

$$i^*_{LVM}(u;k) = \sum_{\alpha=1}^{n} \lambda_\alpha \cdot i(u_\alpha;k) + \left[1 - \sum_{\alpha=1}^{n} \lambda_\alpha\right] \cdot p(k|y_{SCI}, y_{GW}) \quad [10]$$

where, $i^*_{LVM}(u;k)$ are the estimated local probabilities of presence/absence of contamination to be used for simulation, n is the number of data in the search neighbourhood, λ_α are kriging weights, $i(u_\alpha;k)$ are the local indicator data, and $p(k|y_{SCI}, y_{GW})$ is the conditional probability obtained previously.

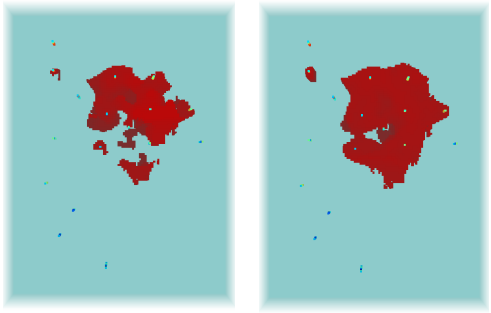

Figure 8. Planar view of two corresponding 3D realizations obtained using SIS-LVM (left) and BU (right)

Bayesian updating is one of the simplest forms of indicator co-simulation. At each location along the random path, indicator kriging is used to estimate the probability of presence/absence of contamination conditioned to local hard data alone ($i^*_{SK}(u;k)$). Then, Bayesian updating updates the probabilities as follows:

$$i^*_{BU}(u;k) = i^*_{SK}(u;k) \cdot \frac{p(k|y_{SCI}, y_{GW})}{p_k} \cdot C \quad [11]$$

where, $i^*_{BU}(u;k)$ are the estimated local probabilities of presence/absence of contamination, p_k is the global probability of absence/presence of contamination, and C is the normalizing constant to ensure that the sum of final probabilities is 1.0.

3.6 Clipping the geostatistical realizations

Following a 'cookie-cutter' approach (Deutsch 2002), the 3D geostatistical realizations obtained by SIS are clipped by 2D realizations of areal extent obtained from the DF approach. This clipping is implemented in a Bayesian framework and ensures the stationarity of the indicator hard data. Two of these 3D clipped realizations obtained from SIS-LVM and BU analyses are shown in Figure 8.

4 CROSS-VALIDATION

Cross-validation techniques are adapted to categorical variables to check the probabilistic prediction of the geostatistical approaches used. Cross-validation with and without conditional probabilities illustrate the value of inclusion of secondary data sources. There are two ways of implementing cross-validation in presence of limited well data: (1) removing each sample and all other samples from the same well, or (2) removing each sample only, while keeping all other samples from the same well. The first option is pessimistic, especially when there only a few wells. The second option is too optimistic. In this work, the first approach is implemented to provide a 'lower bound' on the likely goodness of the prediction, and to evaluate the added value of incorporating the available secondary information.

Table 3. Measures of closeness and percentage improvement, considering secondary data only

	K = 0		K = 1	
	p^0 = 0.733		p^1 = 0.267	
	closeness	**% Imp.**	**closeness**	**% Imp.**
SCI	0.737	0.52	0.283	5.92
GWE	0.757	3.52	0.326	21.94
PR	0.759	3.54	0.344	28.88

According to Deutsch (1999), a quantitative measure of 'closeness' to true categories (presence or absence of contamination) can be summarized by:

$$C_k = E\{p(u_\alpha;k) | \text{true} = k\} \quad , \quad k = 0,1 \quad [12]$$

which may be interpreted as the average predicted probability of the true categories. With no primary or secondary data, the closeness measure will equal the global proportions. The measure of 'percent improvement' over the no-data case is expressed by:

$$C_k^{rel} = \frac{C_k - p_k}{p_k} \quad , \quad k = 0,1 \quad [13]$$

The third measure of goodness that is presented in this work is a measure of 'accuracy'. As we deal with a binary case, at every cross-validation location four cases can be considered: (1) the location is truly contaminated and has been correctly identified, (2) the location is contaminated, but predicted to be clean, (3) the location is clean and predicted to be clean, and (4) the location is clean, but predicted to be contaminated. Cases (1) and (3) are plausible and cases (2) and (4) are not. A measure of 'accuracy' can be defined as:

$$A^{rel} = \frac{M - M_R}{1 - M_R} \quad [14]$$

$$M = \frac{\sum_{i=1}^{N} \left(p_i^{11} + p_i^{00} - p_i^{10} - p_i^{01}\right)}{N} \quad , \quad M_R = p^1 p^1 + p^0 p^0 - 2p^1 p^0$$

in which, N is the number of wells, p_i^{11}, p_i^{10}, p_i^{00} and p_i^{01} are proportions corresponding to the cases 1 to 4, explained earlier, and p^0, p^1 are the global proportions of the two categories. M is the global measure of plausibility. Its upper bound is 1.0, which is obtained in ideal case of correct prediction at all cross-validation locations. Its lower bound is M_R, which corresponds to no-data case.

Table 3 shows some improvements in the predictions, using secondary data only. It can be observed that

incorporating secondary data (in particular groundwater elevation data) considerably improves the prediction of contaminated locations, even before incorporating the hard data. Tables 4, 5 and 6 show the cross-validation results, using the indicator hard data (T-UVIF) as well as secondary soft information such as SCI data, groundwater elevation data (GWE) and their combination with the permanence of ratios (PR) approach. Results from sequential indicator simulation with no secondary data show slight improvement over the global proportions. In all cases, inclusion of secondary data improves the predictive ability. As it was expected, inclusion of secondary data sources improved the prediction ability in almost all cases. Bayesian updating (BU) approach does significantly better than SIS with no secondary data and SIS-LVM in prediction of contaminated locations. However, it seems it slightly over-estimates the presence of contamination and the results tend to be conservative. SIS-LVM also improves the prediction ability for both contaminated and uncontaminated locations.

Table 4. Measure of closeness, accounting for indicator hard data and soft secondary information

Measure of closeness	K = 0			K = 1		
	SIS	LVM	BU	SIS	LVM	BU
no secondary data	**0.7532**	-	-	**0.2801**	-	-
SCI	-	**0.7553**	**0.657**	-	**0.2985**	**0.3989**
GWE	-	**0.7345**	**0.6793**	-	**0.3388**	**0.6633**
PR	-	**0.7638**	**0.6933**	-	**0.3898**	**0.6878**

Table 5. Improvement over global proportions, accounting for hard data and soft secondary information

Percentage improvement	K = 0			K = 1		
	SIS	LVM	BU	SIS	LVM	BU
no secondary data	**2.76**	-	-	**4.86**	-	-
SCI	-	**3.04**	**-10.35**	-	**11.81**	**49.4**
GWE	-	**0.21**	**-7.32**	-	**26.9**	**148.45**
PR	-	**4.2**	**-5.42**	-	**45.99**	**157.6**

Table 6. Relative measure of accuracy, accounting for hard data and secondary data from different sources

Accuracy (%)	SIS	LVM	BU
no secondary data	**4.62**	-	-
SCI	-	**6.28**	**-5.28**
GWE	-	**5.14**	**16.94**
PR	-	**14.01**	**21.23**

5 MONTE CARLO SIMULATIONS (MCS)

To study the effect of uncertain source geometry on downstream concentrations, Monte Carlo Simulations are performed. A moderately heterogeneous synthetic aquifer is considered (figure 9) and fate and transport of dissolved BTEX under uncertain source condition (uncertain geometry and mass transfer rate), uncertain transmissivity field, and uncertain biodegradation rate constant is simulated. The modeling domain is 300m × 160m and modeling cell sizes are 1m × 1m, and a hydraulic gradient of 0.01 is imposed to the synthetic aquifer. The partial differential equation describing the fate and transport of BTEX can be expressed as (Chapelle et al. 2003 with some modifications):

$$\frac{\partial(\theta C)}{\partial t} = \frac{\partial}{\partial x_i}\left(\theta D_{ij}\frac{\partial C}{\partial x_j}\right) - \frac{\partial}{\partial x_i}(\theta v_i C) + \theta R^{NAPL} + R_n \qquad [15]$$

where, C is dissolved concentration of BTEX, θ is effective porosity, x_i is the distance along the respective Cartesian coordinate, D_{ij} is the hydrodynamic dispersion coefficient tensor, v_i is pore water velocity, R^{NAPL} is NAPL dissolution term and R_n is the chemical reaction term. According to Imhoff et al. (1993), the rate of mass transfer between NAPL and water can be expressed as:

$$R^{NAPL} = \max[0, k^{NAPL}(C_s^{eq} - C_s)] \qquad [16]$$

in which, k^{NAPL} is mass transfer rate coefficient, C_s^{eq} is NAPL-water equilibrium concentration and C_s is the actual aqueous concentration. In this work, the chemical reaction term is assumed to be a first-order irreversible rate reaction given by $(-\theta\lambda C)$. This is a reasonable assumption for dissolved BTEX undergoing biodegradation. As part of this work a transport simulator is developed to solve the Equation 15. The method of characteristics has been applied to develop the simulator.

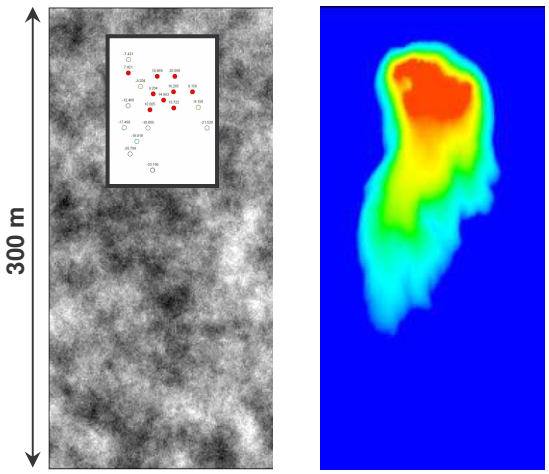

Figure 9. Heterogeneous aquifer and location of suspected source zone (left), and simulated BTEX plume in steady-state condition (right)

The aquifer material is considered to be heterogeneous with a log-normal transmissivity distribution. The mean of the log-normal transmissivity distribution (in natural logarithmic units) is -11.12 (and it is assumed known in Monet Carlo Simulations) and the standard deviation equals to 1.0. The correlation length of the synthetic transmissivity field is L_y = 25m. As L_y is considerably larger than the modeling cell sizes, longitudinal and transverse dispersivities are set to small constant values of 1.0 m and 0.2 m, respectively. In this work, 100 realizations of transmissivity filed conditioned to transmissivity measurements and head observations are generated by sequential Gaussian simulations (Deutsch and Journel 1997). The geometry of the source zone is considered to be uncertain and 100 equi-probabale realizations are generated using the DF approach presented above. k^{NAPL} and first order rate constant λ are also considered to be uncertain. k^{NAPL} has a uniform statistical distribution with a minimum of 0.1 and maximum of 0.6, and λ is log-normally distributed with a mean of 0.0031 day^{-1} and standard deviation (natural-log) of 0.16.

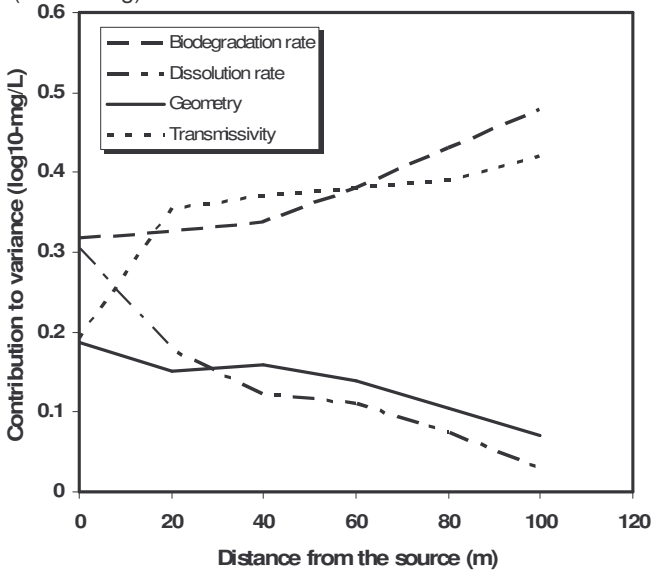

Figure 10. Contribution of uncertain parameters in propagation of uncertainty in average BTEX concentration across different transects downstream of the source area

Figure 10 shows the contribution of each of the uncertain input parameters in propagation of uncertainty in downstream concentrations across different transects. It is observed that for a moderately heterogeneous aquifer the contribution of uncertain biodegradation rate and transmissivity distribution are more significant. Nevertheless, uncertain source zone geometry is also important and accounts for 15 to 20 percent of overall variability at locations close to the source area.

6 CONCLUSIONS

A two-step geostatistical approach was presented to model three-dimensional distribution of residual NAPL, while accounting for secondary sources of information such as soil texture and groundwater elevation. The performance of the proposed methodology was evaluated by cross-validation and the value of secondary data sources and their combination in improving the predictive ability was assessed. Assumption of conditional independence (permanence of ratios) was made to integrate the secondary data. Sequential indicator simulation with locally varying mean as well as Bayesian updating approaches were used to combine the prior probability map with conditional probabilities obtained from secondary data. Results from a set of MCS shows that the uncertainty in source geometry accounts for 15 to 20 percent of overall uncertainty in downstream concentrations at locations closer to source area.

AKNOWLEDGEMENT

The first author would like to acknowledge the Alberta Ingenuity Fund for providing support for this research.

REFERENCES

Brusseau, M.L., Zhang, Z., Nelson, N.T., Cain, R.B., Tick, G.R. and Oostrom M. 2002. Dissolution of Nonuniformly Distributed Immiscible Liquid: Intermediate-Scale Experiments and Mathematical Modeling. *Envir Sci. and Tech.* 36: 1033 – 1041.

Chapelle, F.H., Widdowson, M.A., Brauner, J.S., Mendez, E. and Casey, C.C. 2003. Methodology for estimating Times of Remediation Associated with Monitored Natural Attenuation. *Water-Resources Investigations Report 034057*, US Geological Survey, Columbia, SC.

Deutsch, C.V. and Journel, A.G. 1997. *GSLIB: Geostatistical Software Library and User's Guide*, Oxford University Press, New York, NY.

Deutsch, C.V. 1999. A Short Note on Cross Validation of Facies Simulation Methods. *In Report 1, Centre for Computational Geostatistics*, Edmonton,AB, Canada.

Deutsch, C. V. 2002. *Geostatistical Reservoir Modeling,* Oxford University Press, New York, NY.

Deutsch, C.V. 2006. A Sequential Indicator Simulation program for Categorical Variables with Point and Block Data: BLOCKSIS. *Computers and Geosciences*, 32(10): 1669 – 1681.

Goovaerts, P. 1997. *Geostatistics for Natural Resources Evaluation*, Oxford University Press, New York, NY.

Imhoff, P.T., Jaffe, P.R. and Pinder, G.F. 1993. An experimental study of complete dissolution of a nonaqueous phase liquid in saturated porous media. *Water Resources Research*, 30(2): 307 - 320.

Journel, A. G. 2002. Combining knowledge from diverse sources: An alternative to traditional data independence hypotheses. *Math. Geol.*: 34, 573-596.

Rojas-Avellaneda, D. and Silvan_Cardenas, J.L. 2006. Performance of geostatistical interpolation methods for modeling sample data with non-stationary mean, *Stochastic Env. Res. and Risk Asses.* 20(6), 455-467.

Zhang, Z. and Tumay, M.T. 2003. Non-Traditional Approaches in Soil Classification Derived from the Cone Penetration Test. *In Probabilistic Site Characterization at the National Geotechnical Experimentation Sites*. ASCE, Reston, V.A.