Data Integration for Characterization of Spatial Distribution of Residual Hydrocarbon at a Contaminated Site

Amir H Hosseini

Department of Civil and Environmental Engineering
University of Alberta

In groundwater contamination scenarios associated with petroleum products, residual non-aqueous phase liquids (NAPLs) create a long-term source of contamination. In spite of sophisticated models proposed to quantify the mass transfer between NAPLs and aqueous phase in a laboratory setting, little attention has been paid to the spatial distribution of residual NAPL in a field-scale. In this paper, a two-step geostatistical approach is proposed to characterize the distribution of residual NAPL contamination in a 3D space. As the first step, multiple secondary data sources are combined and used in generating multiple 3D geostatistical realizations of presence/absence of contamination. In the second step, the generated realizations are 'clipped' by 2D realizations of areal extent obtained from a distance function-based approach. A cross-validation exercise is then implemented to show the value of secondary data in improving the prediction ability of the proposed methodology and its overall performance.

Introduction

Many groundwater contamination incidents begin with the release of essentially immiscible fluids into the subsurface environment. The immiscible fluids, termed LNAPLs (Light Non-Aqueous Phase Liquids), are typically produced, stored and distributed as gasoline, diesel, heavy fuel and lubricating oils. The characteristics of these products in conjunction with their geologic and hydrogeologic conditions at a contaminated site are the primary factors that influence the movement and distribution of mobile and residual LNAPL in the subsurface.

When oil (LNAPL) is accidentally released, it migrates vertically and laterally under the gravity and capillary forces. When the volume of the release is sufficient, the LNAPL will migrate through the unsaturated zone to the capillary fringe and the water table (figure 1). Due to capillary forces, some LNAPL is always retained in the soil pores as 'residual' or 'immobile' NAPL. In fact, LNAPL coexists with water (and air) in the soil pores. LNAPL saturations are always less than 100 percent but may range from as little as 5 percent to over 70 percent (figure 2). As the remaining 'mobile' LNAPL continues to migrate through the subsurface, the volume of mobile product decreases as NAPL becomes trapped as isolated droplets within the soil pore network. Thus, LNAPL plumes are 'spatially self-limiting', unless continually supplied from an ongoing release (API 2004). While migrating through the subsurface, LNAPL is affected by the heterogeneous nature of the soil strata: slight differences in soil texture may promote preferential pathways within the aquifer horizontally and vertically. Also, LNAPL is significantly influenced by vertical fluctuations in the water table. These fluctuations enhance the development of the residual LNAPL. The residual NAPL is almost impossible to be removed and creates a long-term source of pollution as it partitions slowly into the aqueous and vapor phases.

Despite numerous studies conducted to quantify the mass transfer (geochemical partitioning) between NAPLs and aqueous phase, characterization of spatial distribution of 'residual LNAPL' has gone unnoticed for a large part. Spatial distribution of 'residual NAPL' is particularly important as an input for contaminant fate and transport models. A literature survey in the area of fate and transport modeling for contaminants associated with petroleum products indicates that almost always 'over-simplifying' assumptions have been adopted (Waddill and Widdowson, 1997).

In this paper, a two-stage geostatistical approach is proposed to delineate the space of uncertainty associated with distribution of residual NAPL in a contaminated aquifer. The proposed methodology is presented in the form of a case study for a hydrocarbon impacted site located at west-central Alberta.

As the first step, data from multiple data sources such as soil texture and groundwater surface elevation are combined with the assumptions of full data independence and conditional independence (permanence of ratios). This gives a 3D map for conditional distribution of presence/absence of contamination conditioned to soil texture and groundwater surface elevation. In a sequential indicator simulation (SIS) context, this conditional distribution is then combined with prior probability map to build a 3-dimentional updated posterior probability map. In this work, indicator hard data as well as soil texture data are originated from Ultra-Violet Induced Cone Penetration Testing (CPT-UVIF) and groundwater elevation data are obtained from 23 piezometers installed at the contaminated site.

Primary Hard Data: Truncated UVIF Readings

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 zones impacted by aromatic hydrocarbons. It records the mechanical responses of the soil at the same scale as it records the UVIF responses. The UVIF responses can be only reliably used as a screening tool to identify contamination by LNAPLs. In other words, UVIF response has a complex relationship with LNAPL concentration and can easily be incorporated into geostatistical modeling. In this work, instead, a categorical variable (T-UVIF) is introduced to represent the presence or absence of contamination based on the UVIF responses:

$$i(\mathbf{u}_\alpha; k) = \begin{cases} 1, & \text{if LNAPL is present } (k = 1) \text{ at location } \mathbf{u}_\alpha \\ 0, & \text{otherwise } (k = 0) \end{cases} \tag{1}$$

Figure 3 shows the location map for T-UVIF data. The global proportions for presence and absence of contamination are 0.733 and 0.267, respectively.

Secondary Soft Data: Soil Texture

One of the most common aquifer conditions influencing LNAPL movement is soil heterogeneities. Differences in soil properties will produce preferential flow. In particular, mobile LNAPL will tend to migrate in more permeable and porous soils. Thus, the distribution of the LNAPL (mobile and residual) will generally correspond to the distribution of the permeable units; and a relationship between the soil texture and presence of residual NAPL is expected. In this study, data from cone penetration testing (CPT) has been used to model the geological structure at the site. CPT continuously records the mechanical responses of in-situ soil in a high resolution fashion. It also records the presence or absence of aromatic hydrocarbons (LNAPLs) at the same scale as mechanical parameters (isotopic sampling). Following the methodology introduced by Zhang and Tumay (2003), Soil Classification Index (SCI) can be calculated at every data location. SCI is considered to be well-correlated with effective porosity. Figure (4-a) shows the histogram for SCI data. The presence/absence of contamination was calibrated against SCI data and a calibration table was established (figure 4-b, table 1). Table 1 depicts a positive correlation between probability of presence of contamination and SCI, which is in turn related to effective porosity.

Table 1: calibration of presence/absence of contamination against Soil Classification Index

		$p(k=1 \mid y_{SCI})$	$p(k=0 \mid y_{SCI})$
SCI –class (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

In order to generate the 3D map of conditional probabilities ($p(k|y_{SCI})$), 100 realizations of SCI field were simulated by Sequential Gaussian Simulation (Deutsch and Journel 1997) on a $120\times160\times56$ grid. The equal-sized cell dimensions were $0.5\text{m}\times0.5\text{m}\times0.25\text{m}$. Appropriate conditional probabilities ($p(k|y_{SCI})$) were assigned to each cell in every realization and then averaged over all realizations.

Secondary Soft Data: Location Relative to Groundwater Table

The vertical movement of the groundwater table affects the volume of mobile and residual LNAPL. Given mobile LNAPL on the water table, a rise in the water table causes the hydrocarbon to migrate upward as water displaces LNAPL from the pore space. As water fills the pore network, LNAPL becomes trapped as droplets of LNAPL (condition of low saturation in figure 2). Because LNAPL droplets are isolated, they remain trapped as distinct islands of LNAPL with the saturated pore network. The droplets remain suspended in the network until the water table elevation drops. Lowering of the water table enables the oil drain from the pore network. During drainage, droplets of oil remain within the pore interfaces leaving residual oil within the unsaturated zone. The resultant vertical movement of the water table produces a residual 'smear zone' within the saturated and unsaturated zones (figure 5). In order to account for the effects of groundwater table fluctuations, an additional parameter, normalized elevation, is introduced as the elevation UVIF data point relative to groundwater table elevation at the same location:

$$Z_{normal} = Z_{UVIF} - Z_{GW} \tag{2}$$

where, Z_{normal} is the normalized elevation at every data point, Z_{UVIF} is the elevation of the data point in the global coordinate system, and Z_{GW} is the elevation of groundwater table at the data location in global coordinate system. The presence/absence of contamination was calibrated against Z_{normal} data and conditional probabilities were calculated. Figure 6 shows the global probabilities of contamination for different classes of normalized elevation. Calibration of absence/presence of contamination has been summarized in table 2.

Table 2: calibration of presence/absence of contamination against normalized elevation

		$p(k=1 \mid y_{GW})$	$p(k=0 \mid y_{GW})$
Z_{normal} – class (y_{GW})	[-4.3m,-0.576 m)	0.212	0.788
	[-0.576 m , 0.387m)	0.294	0.706
	[0.387m ,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

Integration of Secondary Data Sources: Assumption of Full Data Independence

Bayes' law permits the calculation of the conditional probability $p(k|y_{SCI}, y_{GW})$:

$$p(k|y_{SCI}, y_{GW}) = \frac{p(k, y_{SCI}, y_{GW})}{p(y_{SCI}, y_{GW})} \tag{3}$$

with:

$$p(k, y_{SCI}, y_{GW}) = p(k)\times p(y_{SCI}|k)\times p(y_{GW}|k, y_{SCI})$$

The easiest way to combine the single event probabilities is to assume independence of the two data events. This is, however, a strong assumption and should be taken with care. The assumption of data independence states that y_{SCI} and y_{GW} are independent $p(y_{SCI}, y_{GW}) = p(y_{SCI})\times p(y_{GW})$. An additional assumption is required to simplify equation 3, which is conditional independence of events y_{SCI} and y_{GW} given event $k = 0$ or 1, that is:

$$p(y_{SCI}|k, y_{GW}) = p(y_{SCI}|k)$$

$$p(y_{GW}|k, y_{SCI}) = p(y_{GW}|k)$$

These assumptions result in:

$$p(k|y_{SCI}, y_{GW}) = \frac{p(k|y_{SCI}) \times p(k|y_{GW})}{p(k)} \tag{4}$$

Assumption of full data independence is not robust against departures from the assumption of independence (Ortiz 2003). As a first attempt, in this work, the conditional probabilities have been combined with the hypothesis of conditional independence. Figure 7 –a shows the conditional probability of $k = 1$ given data events y_{SCI} and y_{GW}.

Integration of Secondary Data Sources: Assumption of Conditional Independence

A more robust approach is to assume the data are conditionally independent given the primary data event ($k = 0$ or 1). The expression for conditional probability of the primary data event k given the secondary data events y_{SCI} and y_{GW} is:

$$p(k|y_{SCI}, y_{GW}) = \frac{p(k) \times p(y_{SCI}|k) \times p(y_{GW}|k)}{p(y_{SCI}, y_{GW})} \tag{5}$$

in which, the joint probability $p(y_{SCI}, y_{GW})$ is needed. According to Journel (2002) Bayesian analysis goes around this problem by considering ratios of updated probabilities of the type. This results in the expression for the permanence of ratios assumption:

$$\frac{\dfrac{p(\tilde{k}|y_{SCI}, y_{GW})}{p(k|y_{SCI}, y_{GW})}}{\dfrac{p(\tilde{k}|y_{SCI})}{p(k|y_{SCI})}} = \frac{\dfrac{p(\tilde{k}|y_{GW})}{p(k|y_{GW})}}{\dfrac{p(\tilde{k})}{p(k)}} \tag{6}$$

where, the event $\tilde{k}$ represents the complement of the primary data event k. Expression 6 results in the expression for the conditional probability based on the assumption of permanence of ratios (conditional independence):

$$p(k|y_{SCI}, y_{GW}) = \frac{\dfrac{p(\tilde{k})}{p(k)}}{\dfrac{p(\tilde{k})}{p(k)} + \dfrac{p(\tilde{k}|y_{SCI})}{p(k|y_{SCI})} \cdot \dfrac{p(\tilde{k}|y_{GW})}{p(k|y_{GW})}} \tag{7}$$

Figure 7 –b illustrates the conditional probability of $k = 1$ given data events y_{SCI} and y_{GW}, based on the assumption of conditional independence. According to figure 7, the differences in the conditional probabilities obtained based on the assumptions of full data independence and conditional independence are minor.

Integration of Prior Probability Map with the Conditional Probabilities

The next step is combining the conditional probabilities obtained in the previous steps with the prior probability map (conditioned to indicator hard data only). This was achieved by performing Sequential Indicator Simulation (SIS). There are a number of techniques used to constrain SIS to soft secondary data. In this work, two different techniques are used and their results are compared: (1) SIS with a Locally Vary Mean (LVM) and (2) Bayesian Updating (BU).

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

$$i^*_{LVM}(\mathbf{u};k) = \sum_{\alpha=1}^{n} \lambda_\alpha \cdot i(\mathbf{u}_\alpha;k) + \left[1 - \sum_{\alpha=1}^{n} \lambda_\alpha\right] \cdot p(k|y_{SCI}, y_{GW}) \tag{8}$$

where, $i^*_{LVM}(\mathbf{u};k)$, k = 0, 1 are the estimated local probabilities of presence/absence of contamination to be used for simulation, n is the number of local data, λ_α, α = 1, …, n are the weights, $i(\mathbf{u}_\alpha;k)$ are the local indicator data, and $p(k|y_{SCI}, y_{GW})$ is the conditional probability obtained previously. Figure 8 (a) and (b) show the planar view of two 3D realizations obtained by LVM approach.

Bayesian updating is one of the simplest forms of indicator cokriging: 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 ($i^*_{SK}(\mathbf{u};k)$). Then, Bayesian updating modifies or updates the probabilities as follows:

$$i^*_{BU}(\mathbf{u};k) = i^*_{SK}(\mathbf{u};k) \cdot \frac{p(k|y_{SCI}, y_{GW})}{p_k} \cdot C \tag{9}$$

where, $i^*_{BU}(\mathbf{u};k)$, k = 0, 1 are the estimated local probabilities of presence/absence of contamination, p_k , k = 0, 1 is the global probability of absence/presence of contamination, and C is the normalization constant to ensure that the sum of the final probabilities is 1.0 (Deutsch 2002). Figure 8 (c) and (d) show the planar view of two 3D realizations obtained by BU approach.

Clipping the Geostatistical Realizations

The second major step in the proposed methodology is: following a 'cookie-cutter' approach, the 3D geostatistical realizations obtained by SIS are clipped by the 2D realizations of areal extent obtained from Distance Function-based approach. The details of the Distance Function-based approach in quantifying the uncertainty in areal limits can be found in Hosseini and Deutsch (2007). Figure 8 shows the 3D realizations of presence/absence of contamination after being clipped.

Cross-validation of Geostatistical Model

Cross-validation methods are adapted to categorical variables to check the probabilistic prediction of the geostatistical techniques applied in this work. A cross-validation study with and without conditional probabilities from the secondary data illustrates the value of the adopted secondary data. 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 (Deutsch 1999). The first option is pessimistic, especially in presence of only a few wells. The second option is overly optimistic. In this work, the first approach is implemented (1) to provide a 'lower-bound' on the likely 'goodness' of the prediction, and (2) to evaluate the added value of having secondary information.

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

$$C_k = E\{p(\mathbf{u}_\alpha;k) | \text{true} = k\}, k = 1,2 \tag{10}$$

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

$$C_k^{rel} = \frac{C_k - p_k}{p_k}, k = 1,2 \tag{11}$$

The third measure of 'goodness' that is considered in this work is the measure of 'accuracy'. As we deal with a binary case (contaminated or clean), at every cross-validation location four cases can be considered in terms of prediction of the true categories: (1) the location is truly contaminated and this contamination is correctly predicted; (2) the location is contaminated, but it is wrongly predicted to be clean; (3) the location is uncontaminated and is correctly predicted as clean; and (4) the location is clean, but it is wrongly predicted to be contaminated. Cases (1) and (3) are plausible and cases (2) and (4) are not. A measure of 'accuracy' of predictions can be defined as:

$$A^{rel} = \frac{M - M_R}{1 - M_R} \tag{12}$$

with:

$$M = \frac{\sum_{i=1}^{N} \left(p_i^{11} + p_i^{00} - p_i^{10} - p_i^{01} \right)}{N}$$

an

$$M_R = p^1 \cdot p^1 + p^0 \cdot p^0 - 2p^1 \cdot p^0$$

where, N is the number of wells removed and replaced in the cross-validation exercise, p_i^{11}, p_i^{10}, p_i^{00} and p_i^{01} are proportions corresponding to the cases 1 to 4, respectively; and p^1 and p^0 are global proportions of categories 1 (contaminated) and 0 (clean). M is the global measure of plausibility. Its upper bound is 1.0, in the ideal case of correct prediction at all cross-validation locations. Its lower bound is M_R, which corresponds to no-data case. Cross-validation results have been summarized in tables 3 to 6.

Table 3: measure of closeness and percentage improvement over the global probabilities, considering the secondary data only (no primary data used)

	K = 0		**K = 1**	
	Global proportion = 0.733		**Global proportion = 0.267**	
	Closeness	**% improvement**	**Closeness**	**% improvement**
SCI	0.7368	0.52	0.2828	5.92
GW	0.7568	3.25	0.3256	21.94
Full DI	0.7574	3.33	0.3422	28.17
PR	0.7589	3.54	0.3441	28.88

Table 3 shows some improvements in the predictions, using secondary data only. It can be observed that considering secondary data (particularly groundwater elevation data) considerably improves the prediction of contaminated locations, even before incorporating the hard data.

Table 4: measure of closeness, accounting for indicator hard data and secondary data from different data sources.

Closeness measures	**K = 0**			**K = 1**		
	IK	**LVM**	**BU**	**IK**	**LVM**	**BU**
no secondary data	0.7532	-	-	0.2801	-	-
SCI	-	0.7553	0.6570	-	0.2985	0.3989
GW	-	0.7345	0.6793	-	0.3388	0.6633
DI	-	0.7639	0.6917	-	0.3875	0.6834
PR	-	0.7638	0.6933	-	0.3898	0.6878

Table 5: percentage improvement over global proportions, accounting for indicator hard data and secondary data from different data sources.

Percent improvement relative to global probability	K = 0			K = 1		
	IK	LVM	BU	IK	LVM	BU
no secondary data	2.76	-	-	4.86	-	-
SCI	-	3.04	-10.35	-	11.81	49.40
GW	-	0.21	-7.32	-	26.90	148.45
DI	-	4.22	-5.36	-	45.16	155.98
PR	-	4.20	-5.42	-	45.99	157.60

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

Accuracy (%)	IK	LVM	BU
no secondary data	4.62	-	-
SCI	-	6.28	-5.28
GW	-	5.14	16.94
DI	-	13.97	20.63
PR	-	14.01	21.23

Tables 4, 5 and 6 show the cross-validation results, while the indicator hard data (T-UVIF data) and secondary soft information (SCI data, groundwater elevation data and their combination with full data independence and permanence of ratios assumptions) are used. The results of indicator kriging (IK) with no secondary information show slight improvement in predictions over the global proportions. In all cases, inclusion of secondary data improves the predictive ability. Nevertheless, as it was expected (figure 7), results of the analysis with the assumptions of full data independence and permanence of ratios are very close. Bayesian updating (BU) technique does significantly better than IK and locally vary mean (LVM) technique in prediction of contaminated locations. However, it somewhat over-estimates the presence of contamination and its results tend to be ‘conservative’. LVM also improves the predictive ability for prediction of both contaminated and uncontaminated locations.

Conclusions

Numerous studies have dealt with partitioning of residual LNAPL into groundwater. However, quantification of areal and lateral extent of residual LNAPL has gone unnoticed for a large part. In this paper, a two-step geostatistical approach is introduced to model three-dimensional distribution of residual LNAPL, accounting for secondary sources of information such as soil texture and groundwater elevation. The proposed methodology was evaluated by cross-validation and value of the secondary data and their combination in improving the predictive ability was assessed. Assumptions of full data independence and permanence of ratios for integration of secondary information resulted in very similar outcomes. Indicator kriging with locally vary mean and Bayesian updating techniques were used to combine the prior probability map with conditional probabilities obtained from secondary data. LVM approach resulted in some improvement in predictive ability of both contaminated and uncontaminated locations. Bayesian updating showed significant improvement over the global proportions for contaminated locations. But, it underestimated the clean areas.

References

American Petroleum Institute: API Interactive LANPL Guide, Version 2.0. Environmental systems and Technologies Inc, Blacksburg, VA, 2004.

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

Deutsch, C. V.: Geostatistical Reservoir Modeling. Oxford University Press, New York, 376 pp., 2002.

Deutsch, C.V., Journel, A.G.: GSLIB: Geostatistical Software Library and User's Guide, Oxford University Press, New York, 368 pp., 1997.

Hosseini, A.H., Deutsch, C.V.: A Distance Function based algorithm to quantify uncertainty in areal limits. In Report 9, Centre for Computational Geostatistics, Edmonton, AB, Canada, 2007.

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

Ortiz, J. M.: Naïve Bayes classifiers and Permanence of ratios. In Report 5, Centre for Computational Geostatistics, Edmonton, AB, Canada, 2003.

Waddill, D.W., Widdowson, M. A.: SEAM3D – A numerical model for three-dimensional solute transport and sequential electron acceptor-based bioremediation in groundwater. Vicksburg, Miss., U.S. Army Corps of Engineers, 177 pp, 1997.

Zhang, Z., Tumay, M.T.: Non-Traditional Approaches in Soil Classification Derived from the Cone Penetration Test. In E. VanMarcke and G.A. Fenton, editors. Probabilistic Site Characterization at the National Geotechnical Experimentation Sites. American Society for Civil Engineers, Reston, V.A., 2003.

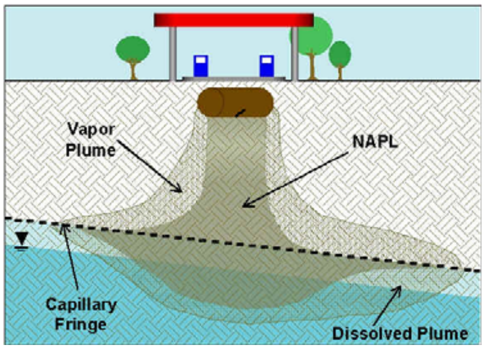

Figure 1: A schematic representation of LNAPL release in the subsurface

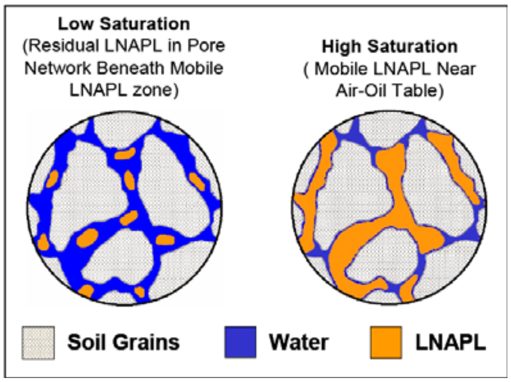
Low Saturation
(Residual LNAPL in Pore Network Beneath Mobile LNAPL zone)
High Saturation
(Mobile LNAPL Near Air-Oil Table)
Soil Grains
Water
LNAPL

Figure 2: Representation of LNAPL saturations in the pore space of saturated zone

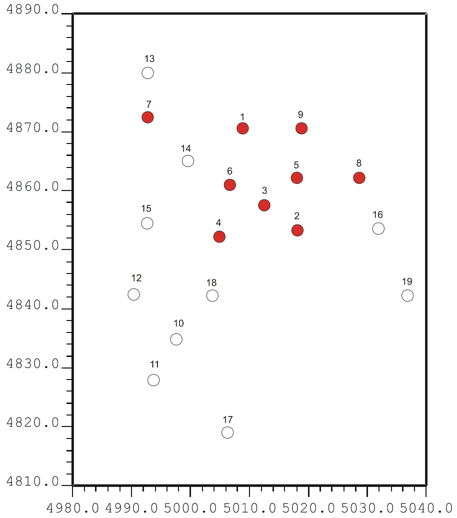

Figure 3: Location map for T-UVIF data. Solid circles represent wells where contamination was detected. No contamination was detected in wells illustrated by hollow circles.

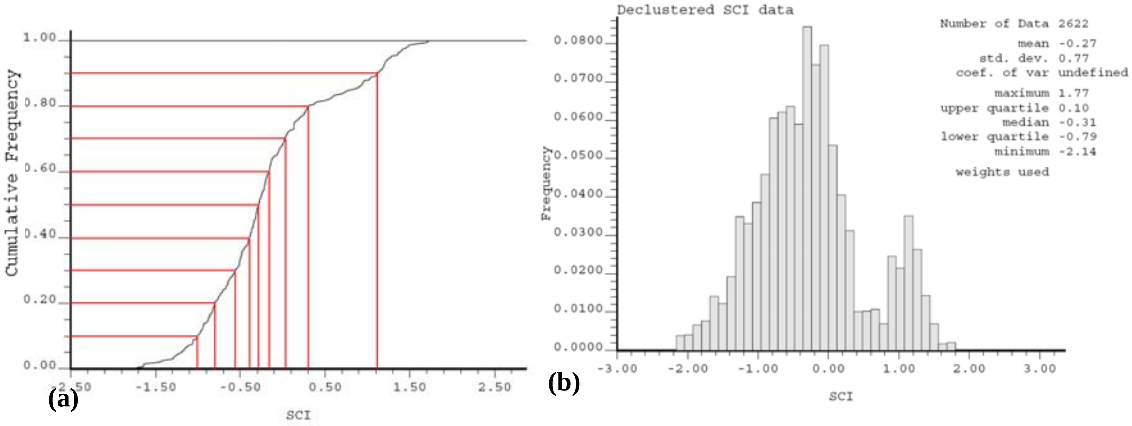

Figure 4: (a) The histogram for SCI data, (b) cumulative histogram of SCI data with 10 classes defined by decile thresholds; there is the same number of data in each threshold.

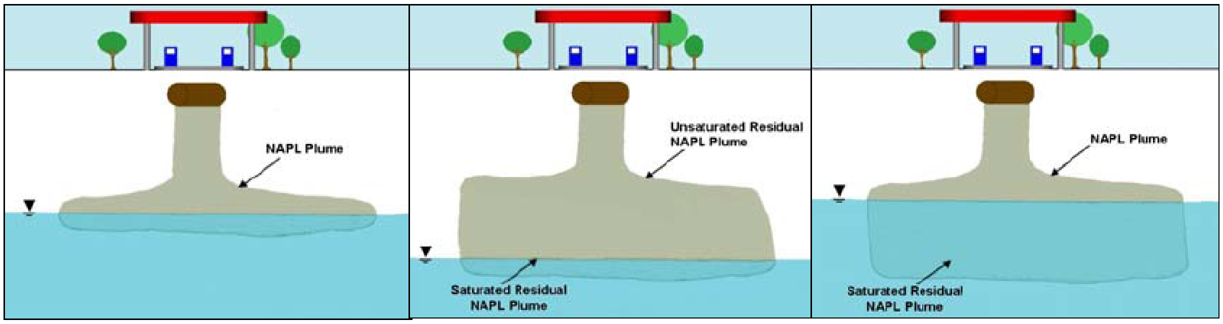

Figure 5: Schematic illustration of LNAPL smear zone created by water table fluctuations.

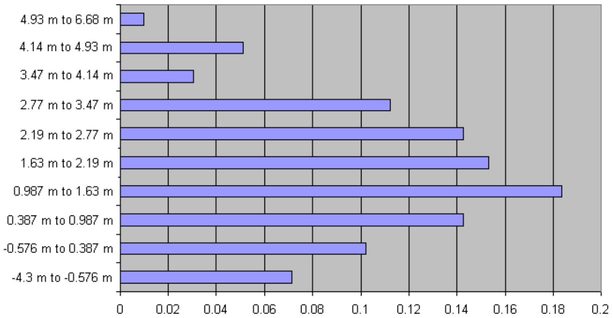

Figure 6: Global proportions of contamination for different classes of normalized elevation. The classes are defined by decile thresholds; there is the same number of data in each threshold.

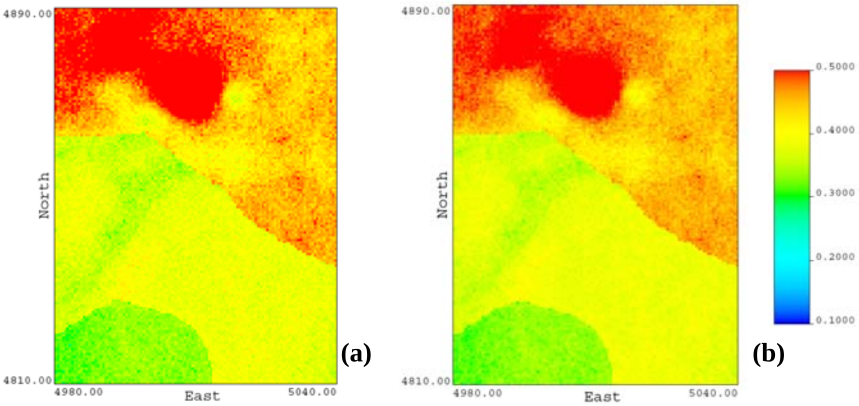

Figure 7: A slice (N_{XY} = 30) from 3D conditional probability map for $p(k=1|y_{SCI},y_{GW})$, (a) based on full data independence, and (b) based on conditional independence (permanence of ratios).

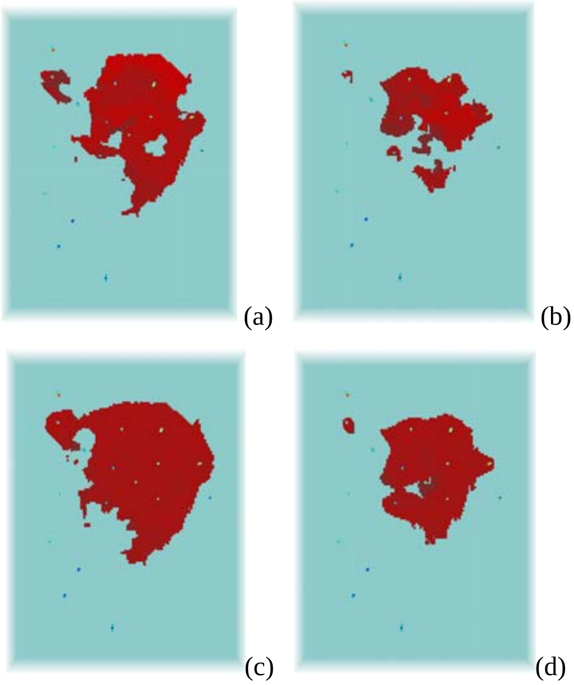

Figure 8: Views of 3D geostatistical realizations of presence/absence of contamination after being clipped. Realizations (a) and (b) are obtained by the LVM technique and realizations (c) and (d) are obtained by the BU technique. The same random number seed was used.