Journal of Hydrology X 13 (2021) 100111

Contents lists available at ScienceDirect

Journal of Hydrology X

journal homepage: www.sciencedirect.com/journal/journal-of-hydrology-x

Research papers

Application of an iterative source localization strategy at a chlorinated solvent site

E. Essouayed, T. Ferré, G. Cohen, N. Guiserix, O. Atteia *

UMR EPOC, Bordeaux-INP, 1 alle, Daguin 33607, France

ARTICLE INFO

Keywords:
Contaminant source localization
Iterative strategy
GLMA
Data Worth
Contaminated site management
Field estimation

ABSTRACT

This study presents an inverse modeling strategy for organic contaminant source localization. The approach infers the hydraulic conductivity field, the dispersivity, and the source zone location. Beginning with initial observed data of contaminant concentration and hydraulic head, the method follows an iterative strategy of adding new observations and revising the source location estimate. Non-linear optimization using the Gauss-Levenberg-Marquardt Algorithm (PEST++) is tested at a real contaminated site. Then a limited number of drilling locations are added, with their positions guided by the Data Worth analysis capabilities of PYEMU. The first phase of PEST++, with PYEMU guidance, followed by addition of monitoring wells provided an initial source location and identified four additional drilling locations. The second phase confirmed the source location, but the estimated hydraulic conductivity field and the Darcy flux were too far from the measured values. The mismatch led to a revised conceptual site model that included two distinct zones, each with a plume emanating from a separate source. A third inverse modelling phase was conducted with the revised site conceptual model. Finally, the source location was compared to results from a Geoprobe@ MiHPT campaign and historical records, confirming both source locations. By merging measurement and modeling in a coupled, iterative framework, two contaminant sources were located through only two drilling campaigns while also reforming the conceptual model of the site.

1. Introduction

The location of contaminated groundwater plumes remains one of the most difficult challenges in contaminant hydrogeology. This task is challenging predominantly due to geologic heterogeneity and uncertainty regarding pollution sources (Atmadja and Bagtzoglou, 2001). Well defined site characteristics (geology and hydrogeology) can facilitate source location, as shown by the success of High-Resolution Site Characterization (HRSC) methods (Barber et al., 2014; McCall et al., 2014; McCall et al., 2017; Meyer et al. 2008; Rosenberg et al. 2021). These strategies and techniques use scale-appropriate measurement and sample density to define contaminant distributions and the physical context of the area with better certainty, supporting faster and more effective site cleanup. This work has been advanced considerably by the capabilities of Geoprobe© MiHPT, to sample the subsurface more rapidly and with higher resolution than standard drilling techniques. However, although is the direct-push technique can be less expensive than drilling in many situations, it still represents a large cost for site remediation. Clearly, there are great advantages to use methods that can reduce the number of samples needed by strategically placing them in locations that carry important information regarding the plume location.

Over the past 30 years, groundwater flow and pollutant transport models have been coupled with inversion methods for source identification (Gorelick et al., 1983; Wagner, 1992; Michalak and Kitanidis, 2004a; Bashi-Azghadi, et al., 2016). Many methods have been tested on synthetic cases, generally relying on concentration data. These methods, compiled in Table 1 and described in more detail in Essouayed et al. (2020), can be classified as: nonlinear optimization; geostatistical; or backward simulation (Bagtzoglou and Atmadja 2005). Unfortunately, few of the proposed methods have been tested on real cases and because of their dependence on knowledge of the hydraulic parameters, or assumptions of homogeneity, they may not be applicable on real sites. A recent exception to this is the study by Hwang et al. (2020) who applied

Abbreviations: GLMA, (Gauss Levenberg Marquardt Algorithm); DW, (data worth); SVD, (Singular Value Decomposition); RMSE, (Root-Mean-Square-Error); N-RMSE, (Normalized-Root-Mean-Square-Error.

* Corresponding author.
E-mail address: olivier.atteia@ensegid.fr (O. Atteia).

https://doi.org/10.1016/j.hydroa.2021.100111
Received 26 June 2021; Received in revised form 13 November 2021; Accepted 17 November 2021
Available online 26 November 2021

Table 1
Existing approaches with their characteristics (Hm: homogeneous, Ht: heterogeneous, K : known, U : unknown).

Method	References	*Kfield*	α_L	*Source*
Linear Optimization	Gorelick et al. (1983)	Hm/K	K	Transient
Non-linear	Wagner (1992)	Hm/K	K	Continuous
Optimization	Mahar and Datta (1997)	Hm/K	K	Continuous
	Mahar and Datta (2000)	Hm/K	K	Transient
	Sun et al. (2006)	Hm/K	K	Transient
	Datta et al. (2009)	Hm/U	K	Transient
	Cao et al. (2019)*	Ht*/K	K	Transient
	Zhang et al. (2015)	Ht/K	K	Transient
	Bashi-Azghadi et al. (2016)	Ht/U	K	Transient
Hybrid method	Aral et al. (2001)	Ht/K	K	Transient
	Singh et al. (2004)	Hm/K	K	Transient
	Ayvaz (2016)	Ht/K	K	Transient
	Xu and Gómez-Hernández (2016)	Ht/K	K	Transient
	Xu and Gómez-Hernández (2018)	Ht/U	K	Transient
Geostatistical method	Snodgrass and Kitanidis (1997)	Hm/K	K	Transient
	Michalak and Kitanidis, (2004a)	Ht/U	K	Transient
	Gzyl, et al. (2014)	Ht/U	K	Transient
	Butera et al. (2013)	Ht/K	K	Transient
Backward	Bagtzoglou et al. (1992)	Ht/K	K	Transient
simulation	Neupauer and Wilson (1999; 2005)	Ht/U	K	Transient
	Michalak and Kitanidis (2004b)	Ht/ Hm/K	K	Transient
	Cupola et al. 2015)	Hm/K	K	Transient
	Hwang et al. (2020)	Hm/K		Transient

*: in the Cao et al. (2019), K value is known at sampling points and several extension of zones of equal values around these points are tested.

a backward transport approach (Neuman et al. 2012) to a real site where the hydraulic conductivity field is supposed to be know precisely. However, as generally the K field has a large uncertainty this approach may indicate a wrong source position at most of the field sites. Another field test has been successful but it required both concentrations and time-lapse electrical resistivity tomography of a tracer injection (Tso et al. 2020). As a result, it is still an outstanding challenge to locate contaminant sources in realistic heterogeneous conditions, with unkown K field, while also minimising sample collection cost.

In practice, an approach with a high computational burden may not be used, even if it could reduce monitoring costs. Therefore, for applications to contaminant transport on real sites, which may require many model runs to account for significant uncertainty, there can be real advantages to employ a parsimonious approach, such as the Gauss-Levenberg-Marquardt-Algorithm method (GLMA). Essouayed et al. (2020) presented the use of GLMA to identify jointly the source location, and the hydraulic and transport properties. However, this study only considered synthetic cases, as a proof of concept of the approach for sites with heterogeneous, unknown values of hydraulic conductivity (K), transverse dispersivity (α_T), and source location for a continuous pollutant release.

The major objective of this study is to locate a contaminant source at a real contaminated site using a new method (Essouayed et al. 2020). To our knowledge, it is the first example of application of an iterative searching strategy at a real site. Due to the site specificities the potential source position is mainly varying along a direction perpendicular to flow. The site has a chlorinated ethenes plume, but the location of the source is uncertain. Very few observations were available at the beginning of the study. The method was tested on a synthetic case study in a previous publication (Essouayed et al. 2020), The present study aims to extend that work by:

- testing if the method can be applied at a real site;
- determining, for this site, how many drilling campaigns are necessary to reach a precise source location;
- analysing the results to see if the estimated parameters, especially the hydraulic conductivity field, are consistent with site knowledge; and
- verifying at the end of the study if the estimated source location was consistent with other information

This last point is key to the study and may seem surprising. In most field studies, historical data and direct-push are used early in the investigation. But our objective was to test the ability of the localization method. Therefore, we withheld this information from the investigators and used it for verification.

Data worth (DW) analysis was used to optimize the location of new measurements to support source location. Data worth analysis cycle to optimize the data collection has been studied widely in hydrogeology (Freeze et al., 1992; James and Freeze,1993; James and Gorelick, 1994; Fu and Gómez-Hernández, 2009; Gates and Kisiel, 1974; Maddock, 1973; Dausman et al., 2010; Hill et al., 2013). In this study, PYEMU (White et al., 2016) is used for measurement optimization. DW and data collection are conducted in an iterative framework: new observations are identified; then the conceptual and numerical models are updated; and the DW is repeated, as necessary. Final testing of the source location is based on comparison with measurements collected after the analyses, namely Geoprobe© MiHPT profiles and detailed historical positions of potential sources.

2. Material and methods

2.1. Site context

2.1.1. General information

The site is located on an industrial park active since 1970 s. Several investigations were carried out on the site but the area for this investigation has not been studied before. The site is polluted by chlorinated ethenes due to historical tanks potentially present inside the building (TCE used for machines degreasing). The source location approach is focused on the zone presented in Fig. 1 with a potential chlorinated ethenes source located below the FF building (a conceptual model is presented in fig. S1 of the supporting information). Before the present study, four monitoring wells of 18–20 m depth were present within the

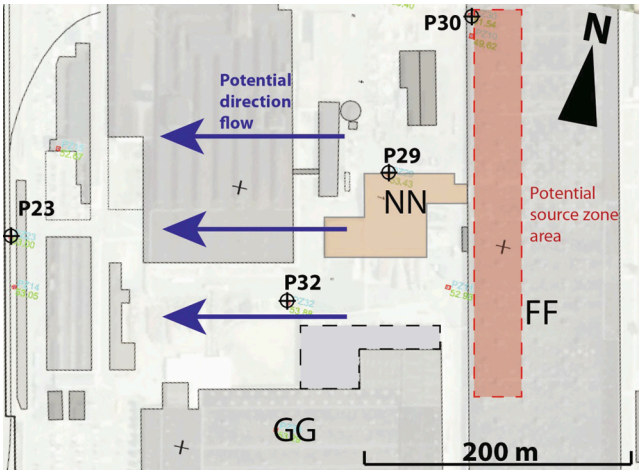

Fig. 1. Map of the industrial site and location of the potentially contaminated building (FF) and the potential contaminated area. At the beginning of this study, 4 monitoring wells are available (P23, P29, P30, P32). Blue arrows indicate the potential flow direction estimated from regional flow. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

zone of investigation (P29, P32, P23 and P30 in Fig. 1).

2.1.2. Geology and hydrogeology

The site is located on a quaternary alluvial formation overlapped on the Upper Cenomanian formation and in the Glauconius Clay of the Lower Cenomanian formation. The main geological facies are:

- filling materials from 0 to 3.5 m deep;
- alluvial formation composed of sand, silex and silty sand from 3.5 to 7 m deep;
- orange-brown coarse sand from 7 to 18 m deep (Upper Cenomanian formation);
- compact grey clay from 18 to 22 m deep (Upper Cenomanian formation); and
- sand from 22 to 30 m deep delineated on the bottom by the Glauconius Clay (Lower Cenomanian formation).

The Cenomanian formation is located in the orange-brown coarse sand from 7 m below ground level (bgl). This layer is 23 m thick and is delineated on the bottom by the Glauconius Clay (lower Cenomanian formation). However, based on the presence of a compact grey clay located at 18 m bgl in piezometer logs available over the whole site, the Cenomaian formation is in fact divided into two aquifers; an upper aquifer with a thickness of 11 m (7–18 m) and a shallow aquifer (22–30 m). That means, that the clay layer limits the movement of the pollutant plume to the shallow aquifer. No chlorinated ethenes concentrations were found in the few piezometers lying in the lower aquifer. No information are available about a potential sorption of solvents on top of the clay layer. The general groundwater flow direction is from the North-East to the South-West

2.2. Site flow and transport model

To cover the area of interest while being able to assign well-defined boundary conditions far from the plume, the domain used for the model extends 980 m in x and 880 m in y (Fig. 3), with 5 m thickness. Water head measurements in existing wells did not show significant vertical gradients, so flow was simulated using a two-dimensional plan view model. Groundwater samples obtained at different depth in the aquifer show small vertical gradient (see fig. S2), suggesting that transport could also be simulated with a 2D approach. To limit computational effort while minimizing impacts on the accuracy of the source location, a variable grid size was used. Specifically, the area of interest (potential contaminated area and monitoring wells in Fig. 1), corresponding to building FF, has a finer grid with a mesh size of 5 × 2 m. The rest of the model has a larger mesh of 20 × 10 m. Flow was modeled with Modflow2005 (Harbaugh 2005).

Monitoring was conducted from July 2017 to August 2018 on P29 and P23 to estimate piezometric temporal variations. The results

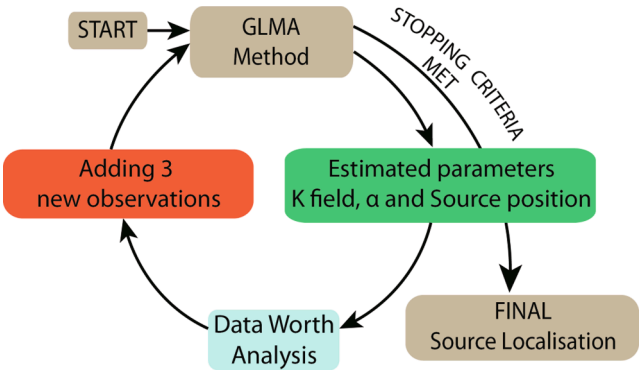

Fig. 2. General diagram of the global approach.

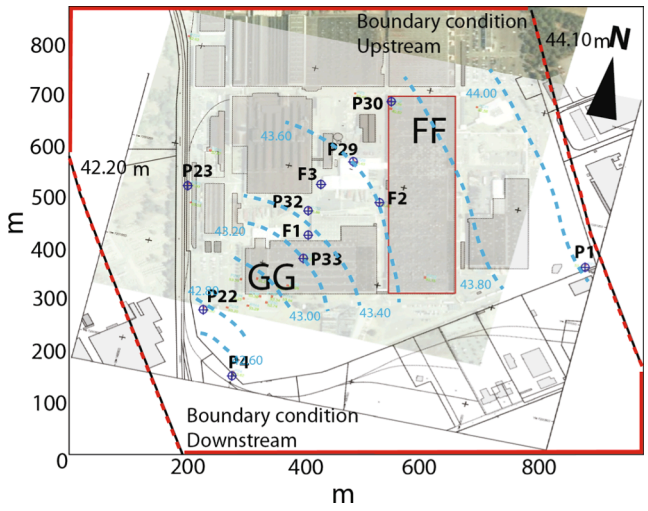

Fig. 3. Initial piezometric map manually drawn in the domain from the 11 wells presented. Model boundary conditions are shown as solid lines for no flow and dashed lines for fixed heads (heads given in the figure)..

showed minimum temporal variations in hydraulic head, allowing for the flow system to be represented as steady state. It was found that after a simulation period of 1400 days, there were no further changes in contaminant concentration within the domain, allowing the plume to be modeled as steady state.

The natural limits represented by a nearby watercourse are located more than 1 km from the site (downstream and upstream). The boundary conditions were defined based on the water levels measured in the monitoring wells located at the downstream and upstream limits of the site (44.10 m in the North-East and 42.20 m in the South-West, Fig. 3). The other boundaries are considered to be coincident with flowlines, represented as no flow boundaries.

2.3. Source location strategy

The global strategy is based on an iterative approach to minimise uncertainties at each source location phase. This strategy is well suited to real-world application, as it can be initiated with observations in only a small number of wells in the plume and then expanded with one or two additional sampling campaigns. The Fig. 2 shows a schematic representation of the strategy. One cycle corresponds to one run of the source location algorithm (GLMA using PEST++ software) and addition of new observations to provide new hydraulic heads (H) and concentrations (C). For each cycle, the method provides estimated parameters (K field, α_L and the source position) including uncertainties on these parameters. The position of the pollution source along the Y-axis, is defined as Ys. The Y axis, aligned with the building's major direction, makes an angle of approximately 30° with the head contours, and is thus the building is oriented nearly perpendicular to the flow direction, which makes the Y axis a good candidate to segregate the source position. Uncertainties linked to Ys are analysed through a DW analysis to identify new drilling locations that are most likely to provide reliable information regarding the source location. In this paper one iteration consists of new drillings, concentrations measurements, GLMA run to estimate the parameters and data worth to find the points for the next iteration.

2.4. GLMA approach

GLMA, implemented through PEST (Doherty, 2015), adjusts selected parameters to reduce the gap between observed and simulated data. In the present study, the adjustable parameters are hydraulic conductivity (K) field (obtained through kriging with pilot points), the longitudinal coordinate of the source (Ys), and the transverse dispersivity (α_T). as the

present work is on a steady plume, the concentrations may vary much more along the Y axis than the X one. This is why it was assumed, as a first step, that the position variation occurred along the Y axis. The number of unknowns (K field and transport parameters) is larger than the number of observations. Therefore, to obtain a stable solution for this under-determined system, Tikhonov regularisation and Singular Value Decomposition (Tonkin and Doherty, 2005) are applied here.

The global approach (Fig. 2) composed of two steps:

(i) an initial optimization to estimate the pilot point parameters for the K field by simulating only the hydraulic head data (H) ;
(ii) a second optimization to estimate all of the parameters based on fitting both the hydraulic head (H) and concentration data (C). The values of the pilot points estimated in Step 1 are not updated in this step.

The values of K and α_L were log10 transformed to stabilize the variation among parameters and the observations were square root transformed to increase the weight on low concentration values in the global objective function. Further details about the parameter settings are given in Essouayed et al. (2020).

2.5. Data worth analysis cycle

The prediction of interest is Ys (y coordinate of the source). Existing data are represented by the contaminant concentration in each cell of the domain once the simulation is completed. Determining the sensitivity of each parameter to the simulated data through GLMA allows consideration of all of the potential sampling locations in the domain that reduce the uncertainty on the prediction of interest (the source position).

Vilhelmsen and Ferré (2018) define the DW as follows:

$$DW = \frac{\sigma^2_{dec}}{\sigma^2_{base}} \quad (1)$$

where σ^2_{dec} and is the variance of the prediction of interest after adding data and σ^2_{base} the variance value of the prediction of interest on the existing data at the beginning of the analysis. DW varies between 0 and 1. For the optimization method, assessment of adding new observations is realised with PyEMU (White et al., 2016). With the Jacobian matrix, PyEMU computes the variance σ^2_{dec} associated with each point (each cell of the model). Then, $(1-DW)$ is calculated to identify zones within which new observations are most likely to reduce the uncertainty of Ys.

Several authors detail the approach for a large number of additional wells (e.g. Xue et al. 2014), however, in practice, it is not feasible to conduct a DW analysis for each new well; it is much more economically reasonable to expect that multiple wells will be sited and contracted together. For this study, the number of additional observations after each iteration has been set to three. A first point (p_1) is chosen in the zone where $(1-DW)$ is the highest. Essouayed et al. (2020) developed a geostatistical approach to selected 2nd and 3rd points. Directional variograms are calculated based on the concentration field (with the simulated plume from GLMA) with the variogram parameters estimated. These parameters are used to construct an ellipse $P(xP, yP, rx, ry)$ centered on the first selected additional observation. The dimensions of the ellipse, rx and ry, correspond to 2/3 of the corresponding range of the two directional variograms (Doherty, 2010). It is assumed that additional measurements within this ellipse will be too highly correlated with p_1. As such, the second point, p_2, is identified as the zone where $(1-DW)$ is highest outside of this ellipse around p_1. Similarly, p_3 is identified as the zone where $(1-DW)$ is the highest outside of ellipses around p_1 and p_2.

2.6. Site parametrization

To represent the site heterogeneity, 95 pilot points were regularly distributed on a relatively loose grid outside the area of interest and with a higher density inside. Kriging with a spherical type variogram and a range of 100 m, and a zero nugget, was chosen to interpolate the hydraulic conductivity among the pilot points.

The parameterization scheme for GMLA for source location is implemented in the same way as in Essouayed et al. (2020), including the hydraulic conductivities represented by the pilot points, K, the Ys coordinate of the source, and the dispersivity, α_T (see Table 2). As the ratio α_T / α_L is kept constant (0.1), varying α_L leads to a proportional variation of α_T which has the major influence on the plume concentrations. The X_s coordinate is considered fixed given the limited information available in the building and the low impact of this parameter on the location of the source.

TCE and *cis*-DCE are the two main chlorinated ethenes present in the aquifer. As at the site no VC was found it is supposed that *cis*-DCE degradation does not occur the TCE + *cis*-DCE sum can thus be considered as a tracer. In the intial stage a few model runs were done to estimate a potential concentration in the source which reached a fixed concentration of 75 μmol.L^{-1}, This is a point source, i.e. its width equal the cell width or 2 m here. The concentration value will be calibrated in the last iteration when enough data will be available. Due to the low concentrations the model only includes dissolved phase transport. Sorption was neglected as the model is run until steady state. The simulation was done with MT3DMS using the TVD scheme for advection.

3. Results

3.1. Iteration 1

3.1.1. Initial state

Because of the small number of existing wells (only 4), a first drilling campaign was conducted before applying the presented approach, adding piezometers F1, F2, and F3. Limited existing information in this area meant that it was possible that there could be a source beneath the building. Initial wells were added to test this possibility. The piezometric map using these points is given in Fig. 3, showing a global flow direction from North-East to South-West.

The measured contaminant concentrations around the FF building are shown on Table 3 and Fig. 4 as the molar sum of TCE + *cis*-DCE. Due

Table 2
Parameters of the GLMA approach used for the site (transformation, increment type and value of the increment refer to the regularization method provided in PEST manual (Doherty 2010).

GROUP OF PARAMETERS		
K	number of pilot points	95
	Variogram range (m)	100
	min - max (m/d)	0.5–450
	Transformation	log
	Increment type	relative
	Value of the increment	0.01
α_T	min - max (m)	0.05–1
	Transformation	log
	Increment type	relative
	Value of the increment	0.01
Ys	min - max (m)	400–710
	Transformation	None
	Increment type	absolute
	Value of the increment	1
OBSERVATION GROUPS		
H	Number	12
	Transformation	Square root
C	Number	6
	Transformation	Square root

Table 3
Concentrations of chlorinated ethenes in Iteration 1 (It1) and Iteration 2 (It2) from May to September.

	cis-DCE (μg L^{-1})	TCE (μg L^{-1})	Σ (μmol L^{-1})						
	May								
(It1)	*Aug.*	*Sept*							
(It2)	*May*								
(It1)	*Aug.*	*Sept*							
(It2)	*May*								
(It1)	*Aug.*	*Sept*							
(It2)									
P29	392	111	351	1728	1550	870	17.24	12.98	10.26
P32	39	28	34.5	292	250	111	2.6	2.2	1.2
P30	–	21	15	–	899	883	–	7.08	6.9
P23	12	<5	<5	175	29	33	1.46	0.22	0.25
F1	18.9	14.46	<5	243	92	82.58	2.05	0.85	0.63
F2	<5	<5	<5	356	384	505	2.72	2.94	3.86
F3	<5	<5	10	395	374	152	3.02	2.86	1.26
F5	–	<5	<5	–	1433	1120	–	8.65	8.55
F6	–	<5	<5	–	150.3	133	–	1.15	1.01
F7	–	<5	2201	–	57.65	145	–	7.67	23.8

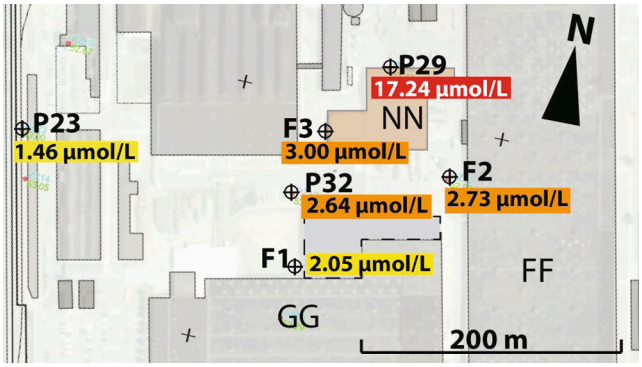

Fig. 4. spatial distribution of the sum of TCE + *cis*-DCE expressed in μmol L^{-1} in Iteration 1.

to limited access to the building, water sampling at P30 was not possible. The highest concentration (17 μmol L^{-1}) was measured at P29, a monitoring well located a priori in the upstream part of the plume. F3, F2 and P32 appear to be located in the middle of the plume with similar concentrations (2–3 μmol L^{-1}). At the downstream site boundary, P23 presents the lowest measured concentration (1.46 μmol L^{-1}).

3.1.2. *Glma*

An initial model was constructed to loosely mimic the measured hydraulic head by a first GLMA optimization. This first step allows the model to have initial value for the *Kpp*, used for the second step of optimization. Then, the GLMA optimization based on the measured concentration and hydraulic head was conducted, leading to an estimated source position at $Ys = 642 \pm 14$ m and a dispersivity of 0.5 ± 0.13 m. The estimated steady state plume based on the best fit parameter values is presented in Fig. 5. The Root-Mean-Square-Error (RMSE) between simulated and observed data for hydraulic head (*H*) and concentration (*C*) are 0.06 m and 1.2 μmol L^{-1}, respectively. For Iteration 1, the coefficients of variation (Normalized RMSE) were 5% and 7.9% for *H* and *C*, respectively (Fig. S3). This indicated acceptable calibration based on the preselected limits of 5% for *H* and 10% for *C*. The number of PEST iteration steps for the flow fitting is generally between 5 and 10 to reach stabilization but this step is quite fast however. For the transport runs the number of PEST iterations to estimate the K field and dispersivity was limited to 30 due to computation time. For the present model, this leads to approximately 2 days of calculations on a 20 core machine.

3.1.3. *Data worth analysis cycle*

The Data worth (DW) analysis was conducted within the plume area.

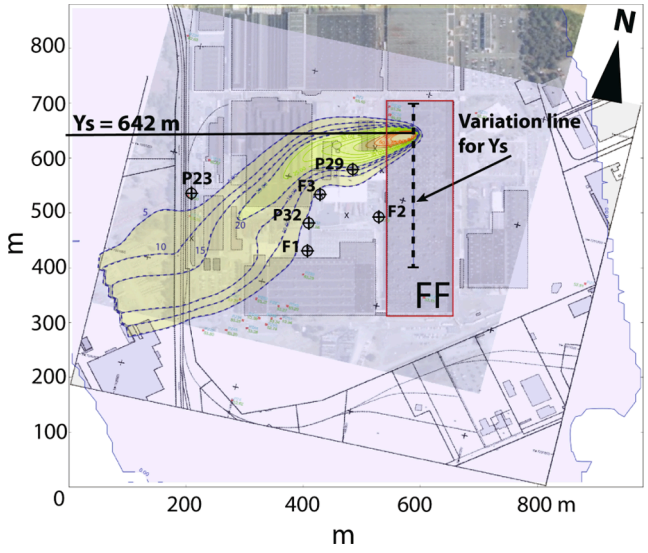

Fig. 5. Location of the source after the GLMA in Iteration 1. The contour presents the concentrations on the plume using the best estimate of the source position (given here as Ys) and dispersivity.

A total of 8720 points (each cell of the model in this area) were added to create the *Ys* uncertainty map. Similar to the strategy shown in Essouayed et al. (2020), two directional variograms were determined from the simulated concentrations following Iteration 1. The ellipses centered on already sampled points have dimensions equal to the variogram ranges, i.e. of 130 m in the x-direction and 55 m in the y-direction.

Inaccessible areas (e.g. beneath buildings) were eliminated from consideration. Piezometer F7 was chosen because it is located in an accessible zone where the uncertainty is the highest (yellow and red zones in Fig. 6a). When identifying the second sampling location, points within an exclusion ellipse around F7 were disqualified. The highest (1-DW) areas in consideration are located northeast of this ellipse. Strictly following the DW protocol would have required that the point with the highest (1-DW) should have been chosen. However, in keeping with the practical nature of this study, an existing piezometer (P30) which could not be sampled during the Iteration 1 and that was located in a comparably high (1-DW) zone, was selected to minimize cost. An ellipse around P30 covers the areas of highest (1-DW). The location with the highest value outside of the first two ellipses is F6. Because an existing unsampled well was used, it was decided to add a fourth sampling location. The accessible areas presenting the highest (1-DW) outside of

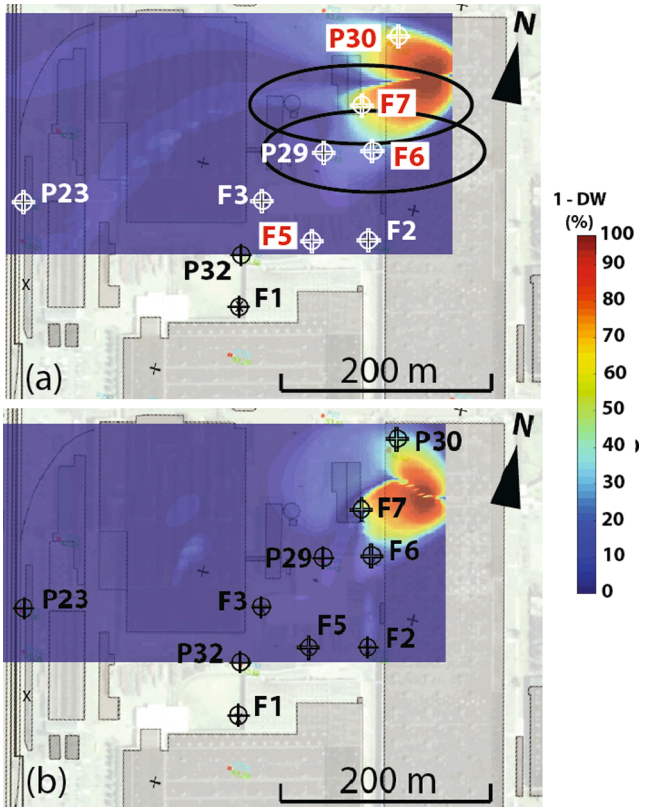

Fig. 6. Distribution of the *Ys* uncertainty (1-DW) in the modeled area. (a) Uncertainties calculated in Iteration 1 with a source located at a Ys = 642 m. The data worth map allows the location of the future monitoring wells (P30, F7, F6 and F5) and the corresponding exclusion ellipses (dashed white lines). (b) Uncertainties calculated in Iteration 2 with a source located at a Ys = 634 m thanks to the new observations. In addition, observe that uncertainties in Iteration 2 are lower than in Iteration 1.

the three ellipses around the first three sampling points led to the choice of F5 as the fourth sampling location.

3.2. Iteration 2

3.2.1. Initial state

Iteration 2 sampling was conducted on September 2018 (one month after drilling) in 10 piezometers (F1, F2, F3, F5, F6, F7, P30, P29, P32 and P23). The results of the chlorinated ethenes concentrations measurements are presented in Table 3. Fig. 7 summarizes the concentrations measured at the beginning of Iteration 2. The maximum combined molar concentration (23.80 μmol L^{-1}) was observed in F7. It appears that F6 (1.01 μmol L^{-1}) is not affected by the plume. It can be seen that F6 is surrounded by locations (F7, F5, and F2) that show evidence of contamination, but are farther downgradient than F6.

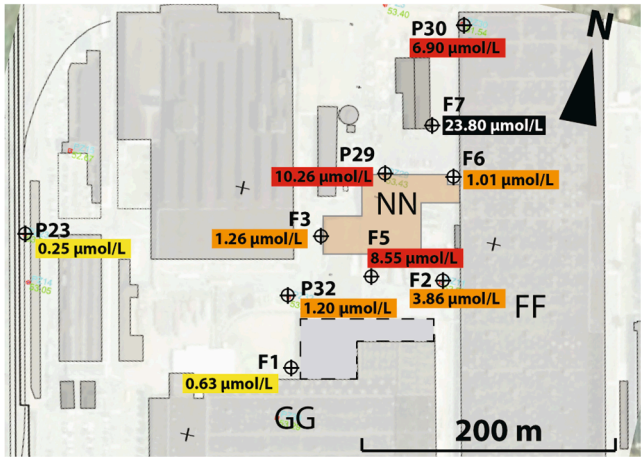

Fig. 7. spatial distribution of the molar sum of chlorinated ethenes measured in Iteration 2. Color linked to concentrations.

3.2.2. GLMA

Following Iteration 1, the inferred values were: $Ys = 642 \pm 14$ m and $\alpha_L = 4.97 \pm 1.27$ m. Optimization based on all of the observations (15 for *H* and 10 for *C*), results in a source allocation of $Ys = 634 \pm 3.66$ m and a dispersivity α_T of 0.51 ± 0.11 m (Fig. 8b). The RMSE of *H* and *C* are 0.08 m and 2.5 μmol L^{-1} (NRMSE - H = 5% and NRMSE - C = 11%), respectively (Fig. S3). This represents considerable reduction in the uncertainty of both parameters.

3.2.3. Data worth analysis cycle

The reduction in uncertainty following each iteration can also be seen in the DW maps. Comparing the area covered by high (1-DW) values after Iteration 1 (Fig. 6) and Iteration 2 (Fig. 6b) shows both a relocation of the areas likely to provide additional information as well as a reduction in the high (1-DW) area.

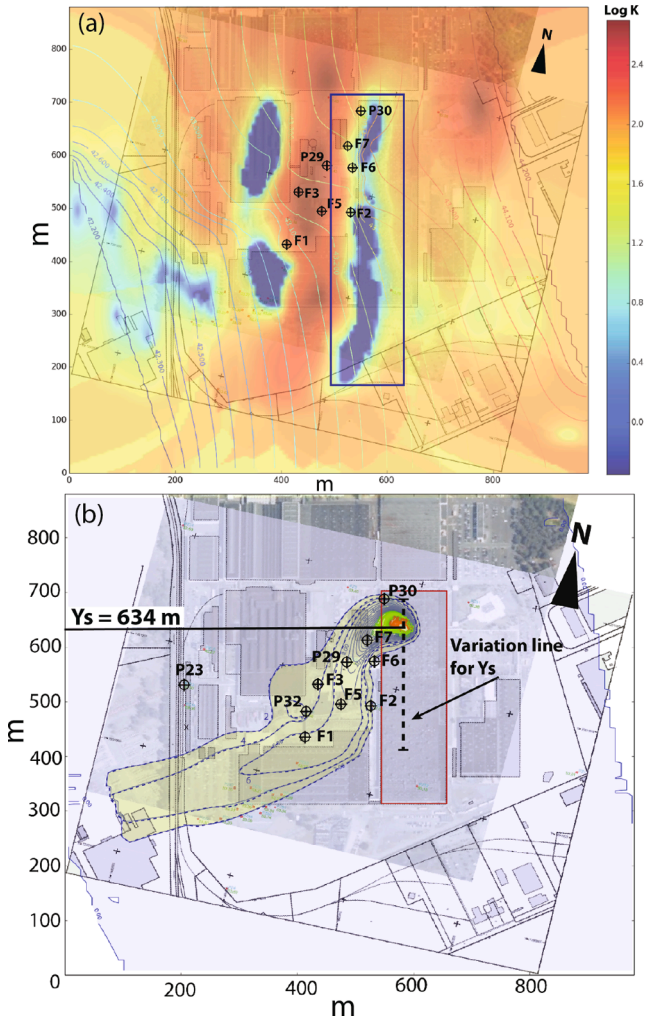

Fig. 8. (a) Spatial distribution of the hydraulic conductivity at the end of Iteration 2 (in $\log_{10}$(m d^{-1}). The anomaly found is framed in blue. (b) Location of the source and plume simulated in Iteration 2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.3. Iteration 3

3.3.1. Initial state – Iteration 3

Hydrogeologic investigations are commonly data-limited. As a result, it is common to collect data that lead to hydrologic 'surprises' (Bredehoeft, 2005) as the example in Iteration 2. The sequential nature of the proposed approach encourages an investigator to reanalyse data as they are collected and to use the updated analysis to plan additional data collection. Indeed, K field estimated by the GMLA model following Iteration 2 indicates areas of very low *K* at the edge of the FF building (Fig. 8). However, this low *K* feature is likely an anomaly that does not represent the real context of the geology and soil found at the site (coarse sand and no impermeable foundation present close to the FF building at such depth). In addition, the Darcy fluxes measured (integrated over the representative depth) using the Direct Velocity Tool (Essouayed et al., 2019) in the field do not agree with the velocities obtained by the model (blue points in Fig. S4). This additional information made it possible to evaluate the origins of the uncertainties encountered in Iteration 2 for source location.

Based on the poor agreement between the inferred K field and knowledge of the site geology, on the disagreement between the model-derived Darcy fluxes obtained and field measurements, it must be considered that the conceptual model of the site was inadequate. Specifically, it was decided to test an alternative model that proposes the existence of a second source zone. In this case, the results suggested that an acceptable model could not be formed based on a single contaminant source. However, as with many real-world investigations, data had already been collected. Therefore, the first step was to assess the possibility of there being more than one source based on the data gathered in Iteration 2.

During the field investigations, measurements of pollutant concentrations and Darcy fluxes in the monitoring wells close to the FF building were conducted using the DVT (Essouayed et al., 2019). This made it possible to evaluate the contaminant mass fluxes passing through each piezometer (Fig. 9a), giving additional information on the plume dynamics.

F3 and F6 wells show quite small fluxes, which suggest that a plume is not crossing this region. Moreover, well F5 contains the highest TCE concentration but no DCE (Fig. 9b), suggesting that it is not possible that the water sampled at this point came from the north (F6 and F7). Such a TCE concentration, with the observed low fluxes zone along F3 and F6, would however be consistent with the presence of a second TCE plume in the south. Two separate plumes are shown as color-filled elliptical areas on Fig. 9a. The results shown on Fig. 9b suggest that the northern plume represents degradation of TCE to *cis*-DCE close to the source, as *cis*-DCE is present over the whole zone, whereas the southern plume is entirely composed of TCE.

3.3.2. GLMA – Iteration 3

In order to integrate this new conceptual model, two sources were manually included in the numerical model. One source was placed at $Ys1 = 620$ m with a concentration of 50 µmol L^{-1} and a second one $Ys2 = 522$ m with an initial source concentration of 40 µmol L^{-1}. Transverse dispersivity was considered to be known with a value of 0.51 m, estimated during the previous GMLA analysis. The pilot points were calibrated approximately before optimization began. Then, the GMLA optimization was done and the results show: (i) a first source, *Ys*1, at a distance equal to 624 ± 8 m and a concentration of 72 µmol L^{-1} and (ii) a second source, *Ys*2, at a distance equal to 533 ± 6 m and a concentration of 17 µmol.L^{-1} (Fig. 10b). The RMSE values are equal to 0.08 m (NRMSE = 4.6%) and 1.9 µmol.L-1 (NRMSE = 8%) for H and C, respectively. The source in the North impacts piezometers F7, P29 and F3 and a source in the South impacts F2, F5 and F1.

As part of this analysis, a new K field was inferred. This new K field shows much better agreement with the site geological characteristics, with smoother spatial variations of logK and the absence of long and straight lines (Fig. 10a). The possible rapid degradation of TCE to *cis*-DCE in the north plume has been validated through reactive transport modelling (not presented here). Owing to the small amount of data on the possible reducing area and the degradation rates we considered that it was not possible to include the reactive model inside the source localization approach, it would have carried too many uncertainty. The reactive transport model was just used to verify that the K field and source position were compatible with the observed *cis*-DCE/TCE ratio (no Vinyl Chloride or Ethene found) at observation points. In addition, the simulated Darcy fluxes are much closer to the measured values than those inferred following the Iteration 2 analysis based on a conceptual model of a single source (red points in Fig. S4).

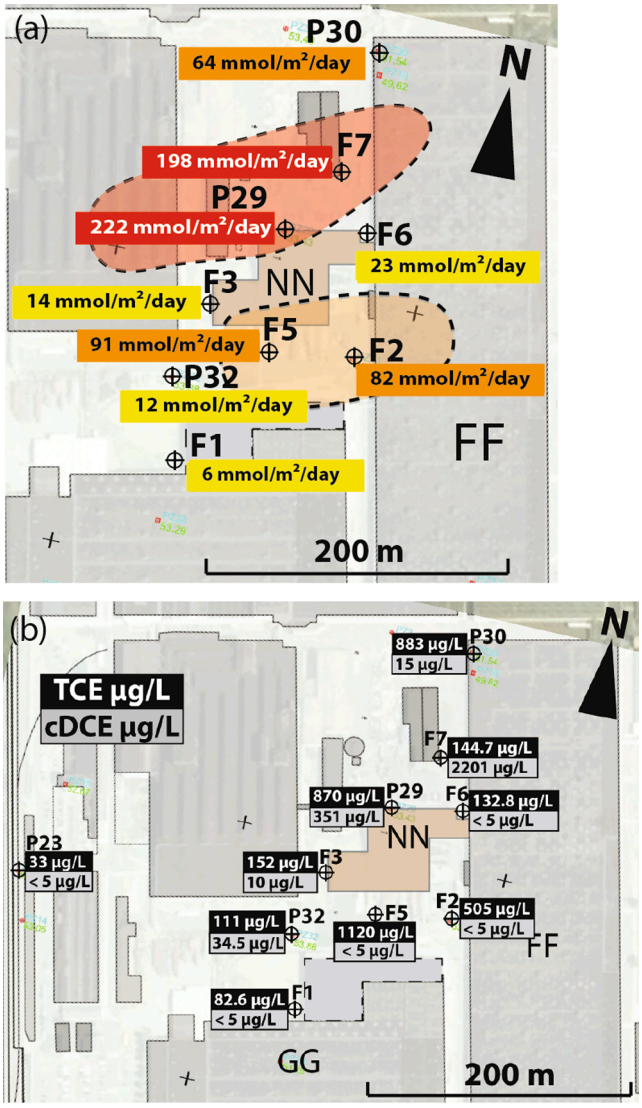

Fig. 9. (a) Distribution of chlorinated ethenes mass flux, the dashed zones represent a potential zone of similar mass fluxes. (b) Concentration of TCE and *cis*-DCE

3.4. Validation

Independently from the source location approach, two strategies were adopted to estimate the source positions:

First, a classical historical review of the storage and main use of chlorinated ethenes in the factory was conducted. This allowed to delineate areas in the factory that were most likely to have been contaminated (Fig. 11). The vertical rectangle (oriented south to north)

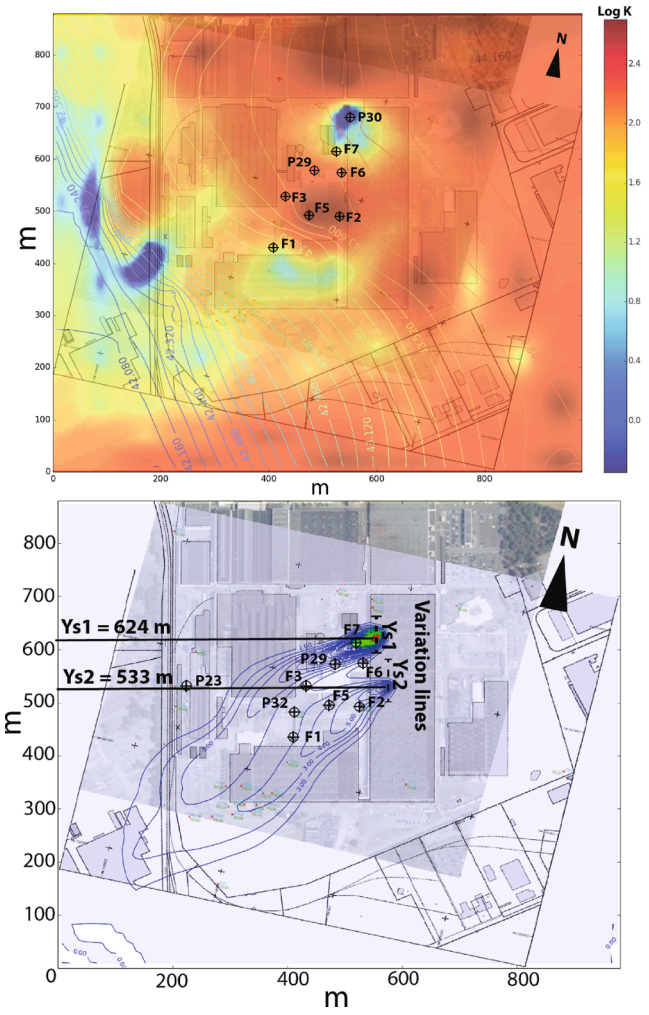

Fig. 10. (a) Spatial distribution of the hydraulic conductivity at the end of Iteration 3 (in $\log_{10}$(m d^{-1})). (b) Location of the 2 sources estimated in Iteration 3.

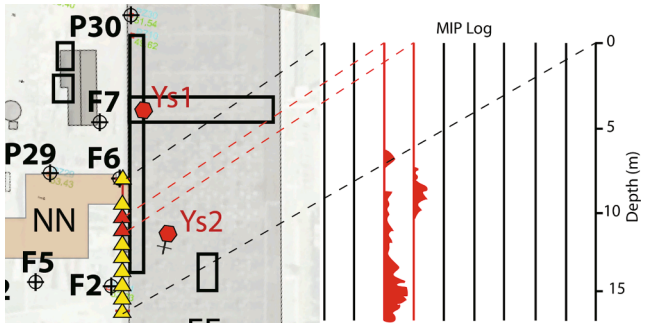

Fig. 11. Position of the sources estimated with the iterative approach (*Ys*1, *Ys*2 octogons). Triangles are the position of the Geoprobe© MiHPT (MiHPT is a combination of MIP (Membrane Interface Probe) for VOC measurement with HPT (High Pressure Tool))transect, with MIP log positioned on the right side and shows impacts in two points from 5 to 15 deep, represented with red triangles. Black rectangles show the positions of the potential (historical analysis) source zones in the factory. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

does not provide more information about a potential location of a source. On the other hand, due to the flow orientation, any leakage in the rectangle oriented in the west to east direction would lead to a plume reaching F7, which thus highly increases the potential presence of a plume around this area. Therefore, the Ys1 source found through the presented approach is likely to come from this area.

The second validation step was done through Geoprobe© MiHPT drilling along a transect (triangles on Fig. 11). The Geoprobe© MiHPT was carried out before the development of the source localization method which explains why only the part of the Ys2 was investigated. Results of this campaign were consulted only at the end of the study, to avoid the knowledge of potential sources during the source localization process.

Chlorinated ethenes were only encountered at two locations (red triangles on Fig. 11) and at quite low concentrations. However, these locations are in good agreement with the inferred position of the *Ys*2 source found through the iterative approach. Unfortunately, the Geoprobe work was completed before the hypothesis of a second source at Ys1, so no results are available yet to test the location of the presence of a source in the north.

4. Discussion and conclusion

The strategy of a contaminant source location was tested on a real contaminated site to evaluate the reliability of the method developed and presented in Essouayed et al. (2020). Due to the limited information available in the chosen zone and in order to perform the method developed in synthetic cases, only one potential source was considered in the FF building at the beginning. Molar sums of TCE and *cis*-DCE and hydraulic head were used as observed values for the first two iterations of the iterative data collection – calibration – data worth analysis cycle.

To our knowledge, this is the first realization of this data collection strategy, including unkown K field, at a real field site. It is certainly the first attempt to combine GLMA and DW. The approach was successful in two ways. First, there was continual reduction of the uncertainty of the location of a source and the transverse dispersivity through the first two iterations of the process. Second, the approach led to clear proposals for additional data collection that are objective and defensible. Third, because the process requires a hydrologist to reevaluate data as it is collected and before new data collection is proposed, the procedure forced to reconsider the conceptual model after Iteration 2. This led to the consideration of complementary data (mass flux measurements) and, ultimately, to the proposal of a second contaminant source. The final round of data collection was guided by this revised conceptual model, leading to a much-improved analysis, consistent with all collected data, following Iteration 3.

The final round of analyses led to two separate sources, which impact F7 and P29 in the North, and F2 and F5 in the South, with different chemical signatures. The two sources were identified at $Ys1 = 623 \pm 8$ m $Ys2 = 532 \pm 6$ m (Fig. 10b). The results of historical records related to potential contaminant sources and to a Geoprobe© MiHPT investigation support the inferred source locations. The use of two sources arose from the poor result of one source at stage 2. Owing to the small amount of data before iteration 2 we think that the 2nd source option could have been tested before this iteration. In general with the proposed method and the small amount of available we cannot suggest any theoretical approach to estimate the number of source zones in practice.

The proposed approach to data collection is designed to reduce the amount of data necessary to optimize specific targets of a field investigation. As shown here, the method does not replace expert judgement, rather it allows an expert to make full use of models to optimize data collection. The approach is limited by the initial knowledge about the site, as embodied in the conceptual model and existing data. But this is the case for any approach to hydrologic field investigations. The real advantage of this modular and sequential approach is that it forces a hydrologic investigation to couple data collection and analyses, continually improving data collection efficiency and effectiveness as understanding of the site improves. This and similar approaches should be used more widely in the industry, eventually completing historical

approaches that rely on extensive data collection at the beginning of a project followed by modeling with little or no use of the model to guide data collection.

This work presents a first test of the method. Application at other sites will be required before it can be determined to be broadly validated. There are interesting subtle points in our results. Specifically, the method is aimed specifically at source location. Therefore, as with all model applications, care must be taken when interpreting other hydrologically interesting results. For example, we show that inverse modeling could match the concentrations if the hydraulic field was allowed to vary. However, for the presented case, we demonstrate that the inferred hydraulic conductivity field was incorrect. Such an error may not be as clear in other cases. This result is, perhaps, to be expected. The limited available data could not uniquely constrain the variable hydraulic conductivity field, the X,Y source location and its concentration, and the dispersivity. Dimensional reduction requires both recognition of the uncertainty of some of the inferred parameters (e.g. K(x,y)) and a willingness to narrow the scope of an investigation to match the available data. In this study, we decided to search for the Y coordinate only, not both the X and Y. This was a practical consideration – the procedure could have been applied to search for the X location, but this would have required drilling inside the factory. The high cost of this data collection meant that this objective was outside the scope of this paper.

Like many site investigations, the presented study is based on a 2D model. Here, we could assume reasonably that vertical variations of concentration were negligible. We believe that the model can be extended to consider site conditions that suggest that flow and transport occur in 3D. But, practically, this would be limited to sites that can be simplified to two or three layers and it would require very detailed information on the plume and geology. The authors would be happy to collaborate with hydrogeologists who would like to extend the method to these more complex conditions.

CRediT authorship contribution statement

E. Essouayed: Conceptualization, Methodology. **T. Ferré:** Writing – review & editing. **G. Cohen:** Writing – review & editing. **N. Guiserix:** Funding acquisition. **O. Atteia:** Conceptualization, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was developed during Elyess ESSOUAYED PhD and supported by INNOVASOL, Bordeaux INP ENSEGID and “EA 4592 Georessources et Environnement”. Data of the developed method is available through Essouayed et al. 2020 with a previous use in synthetic cases.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.hydroa.2021.100111.

References

Aral, M.M., Guan, J., Maslia, M.L., 2001. Identification of contaminant source location and release history in aquifers. J. Hydr. Eng. Am. Soc. Civil Eng. 6 (3), 225–234.
Atmadja, J., Bagtzoglou, A.C., 2001. Pollution source identification in heterogeneous porous media. Water Resour. Res. 37 (8), 2113–2125.
Ayvaz, M.T., 2016. A hybrid simulation–optimization approach for solving the areal groundwater pollution source identification problems. J. Hydrol. (Elsevier) 538, 161–176.
Bagtzoglou, A.C., Dougherty, D.E., Tompson, A.F.B., 1992. Application of particle methods to reliable identification of groundwater pollution sources. Water Resour. Manag. (Springer) 6 (1), 15–23.
Bagtzoglou, A.C., Atmadja, J., 2005. In: The Handbook of Environmental ChemistryWater Pollution. Springer-Verlag, Berlin/Heidelberg, pp. 65–96.
Barber, J., Dyment, S., Pitkin, S., 2014. High Resolution Site Characterisation for Groundwater Short Course. Tech. rep, United States Environmental Protection Agency.
Bashi-Azghadi, S.N., Kerachian, R., Bazargan-Lari, M.R., Nikoo, M.R., 2016. Pollution source identification in groundwater systems: application of regret theory and bayesian networks. Iran. J. Sci. Technol. Tran. Civil Engineering (Springer) 40 (3), 241–249.
Bredehoeft, J., 2005. The conceptualization model problem—surprise. Hydrogeol. J. 13 (1), 37–46.
Butera, I., Tanda, M.G., Zanini, A., 2013. Simultaneous identification of the pollutant release history and the source location in groundwater by means of a geostatistical approach. Stochast. Environ. Res. Risk Assess. (Springer) 27 (5), 1269–1280.
Cao, T., Zeng, X., Wu, J., Wang, D., Sun, Y., Zhu, X., Lin, J., Long, Y., 2019. Groundwater contaminant source identification via Bayesian model selection and uncertainty quantificationIdentification de la source des contaminants dans les eaux souterraines via la sélection d’un modèle bayésien et la quantification de l’incertitudeIdentificación de la fuente de contaminantes de aguas subterráneas mediante la selección del modelo bayesiano y la cuantificación de la incertidumbre 基于贝叶斯模型选择及不确定性量化的地下水污染源识别Identificação de fontes de contaminantes de águas subterrâneas via seleção de modelo Bayesiano e quantificação de incertezas. Hydrogeol. J. 27 (8), 2907–2918.
Cupola, F., Tanda, M.G and Zanini, A. “Laboratory sandbox validation of pollutant source location methods.” Stochastic environmental research and risk assessment (Springer) 29 (2015): 169-182.
Datta, B., Chakrabarty, D., Dhar, A., 2009. Simultaneous identification of unknown groundwater pollution sources and estimation of aquifer parameters. J. Hydrol. (Elsevier) 376 (1-2), 48–57.
Dausman, A.M., Doherty, J., Langevin, C.D., Sukop, M.C., 2010. Quantifying data worth toward reducing predictive uncertainty. Groundwater (Wiley Online Library) 48, 729–740.
Doherty, J. “Calibration and uncertainty analysis for complex environmental models”. (2015).
Doherty, J. “Methodologies and software for PEST-based model predictive uncertainty analysis.” Watermark Numerical Computing, (2010).
Essouayed, E., Verardo, E., Pryet, A., Chassagne, R.L., Atteia, O., 2020. An iterative strategy for contaminant source localisation using GLMA optimization and Data Worth on two synthetic 2D Aquifers. J. Contam. Hydrol. 228, 103554. https://doi.org/10.1016/j.jconhyd.2019.103554.
Essouayed, E., Annable, M.D., Momtbrun, M. and Atteia, O. “An innovative tool for groundwater velocity measurement compared with other tools in laboratory and field tests.” Journal of Hydrology X, (2019): https://doi.org/10.1016/j.hydroa.2018.100008.
Freeze, R.A., James, B., Massmann, J., Sperling, T., Smith, L., 1992. Hydrogeological decision analysis: 4. The concept of data worth and its use in the development of site investigation strategies. Ground Water (Wiley Online Library) 30 (4), 574–588.
Fu, J., Gómez-Hernández, J., 2009. Uncertainty assessment and data worth in groundwater flow and mass transport modeling using a blocking Markov chain Monte Carlo method. J. Hydrol. 364 (3-4), 328–341.
Gates, J.S., Kisiel, C.C., 1974. Worth of additional data to a digital computer model of a groundwater basin. Water Resour. Res. 10 (5), 1031–1038.
Gorelick, S.M., Evans, B., Remson, I., 1983. Identifying sources of groundwater pollution: an optimization approach. Water Resour. Res. 19 (3), 779–790.
Gzyl, G., Zanini, A., Frączek, R., Kura, K., 2014. Contaminant source and release history identification in groundwater: a multi-step approach. J. Contam. Hydrol. 157, 59–72.
Harbaugh, A.W., 2005, MODFLOW-2005, the U.S. Geological Survey modular groundwater model – the Ground-Water Flow Process: U.S. Geological Survey Techniques and Methods 6-A16.
Hill, M.C., Faunt, C.C., Belcher, W.R., Sweetkind, D.S., Tiedeman, C.R., Kavetski, D., 2013. Knowledge, transparency, and refutability in groundwater models, an example from the Death Valley regional groundwater flow system. Phys. Chem. Earth. 64, 105–116.
Hwang, H.-T., Jeen, S.-W., Kaown, D., Lee, S.-S., Sudicky, E.A., Steinmoeller, D.T., Lee, K.-K., 2020. Backward probability model for identifying multiple contaminant source zones under transient variably saturated flow conditions. Water Resour. Res. 56 (4) https://doi.org/10.1029/2019WR025400.
James, B.R., Freeze, R.A., 1993. The worth of data in predicting aquitard continuity in hydrogeological design. Water Resour. Res. 29 (7), 2049–2065.
James, B.R., Gorelick, S.M., 1994. When enough is enough: The worth of monitoring data in aquifer remediation design. Water Resour. Res. 30 (12), 3499–3513.
Maddock, T., 1973. Management model as a tool for studying the worth of data. Water Resour. Res. 9 (2), 270–280.
Mahar, P.S., Datta, B., 2000. Identification of pollution sources in transient groundwater systems. Water Resour. Manage. 14, 209–227.
Mahar, P.S., Datta, B., 1997. Optimal monitoring network and ground-water–pollution source identification. J. Water Resour. Plan. Manag. (Am. Soc. Civil Eng.) 123 (4), 199–207.
McCall, W., Christy, T.M., Pipp, D., Terkelsen, M., Christensen, A., Weber, K., Engelsen, P., 2014. Field Application of the Combined Membrane-Interface Probe and Hydraulic Profiling Tool (MiHpt). Groundwater Monit. Remed. (Wiley Online Library) 34 (2), 85–95.

McCall, W., Christy, T.M., Evald, M.K., 2017. Applying the HPT-GWS for Hydrostratigraphy, Water Quality and Aquifer Recharge Investigations. Groundwater Monitoring & Remediation (Wiley Online Library) 37 (1), 78–91.

Meyer, J.R., Parker, B.L., Cherry, J.A., 2008. Detailed Hydraulic Head Profiles as Essential Data for Defining Hydrogeologic Units in Layered Fractured Sedimentary Rock. Environ. Geol. 56 (1), 27–44. https://doi.org/10.1007/s00254-007-1137-4.

Michalak, A.M., Kitanidis, P.K., 2004a. Application of geostatistical inverse modeling to contaminant source identification at Dover AFB, Delaware. J. Hydraul. Res. (Taylor & Francis Group) 42 (sup1), 9–18.

Michalak, A.M., Kitanidis, P.K., 2004b. Estimation of historical groundwater contaminant distribution using the adjoint state method applied to geostatistical inverse modeling. Water Resour. Res. (Wiley Online Library) 40 (8). https://doi.org/10.1029/2004WR003214.

Neuman, S.P., Xue, L., Ye, M., Lu, D., 2012. Bayesian analysis of data-worth considering model and parameter uncertainties. Adv. Water Resour. 36, 75–85.

Neupauer, R.M., Wilson, J.L., 2005. Backward probability model using multiple observations of contamination to identify groundwater contamination sources at the Massachusetts Military Reservation. Water Resour. Res. (Wiley Online Library) 41 (2). https://doi.org/10.1029/2003WR002974.

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

Rosenberg, L., Broholm, M.M., Tuxen, N., Kerrn-Jespersen, I.H., Lilbæk, G., Bjerg, P.L., 2021. Vertical hydraulic gradient estimation in clay till using MiHPT advanced direct-push technology. Groundwater Monitor. Remed. https://doi.org/10.1111/gwmr.12470.

Singh, R.M., Datta, B., Jain, A., 2004. Identification of unknown groundwater pollution sources using artificial neural networks. J. Water Res. Plan. Manag. (Am. Soc. Civil Eng.) 130 (6), 506–514.

Snodgrass, M.F., Kitanidis, P.K., 1997. A geostatistical approach to contaminant source identification. Water Resour. Res. (Wiley Online Library) 33 (4), 537–546.

Sun, A.Y., Painter, S.L., Wittmeyer, G.W., 2006. A robust approach for iterative contaminant "source location and release history recovery. J. Contam. Hydrol. (Elsevier) 88 (3-4), 181–196.

Tonkin, M.J., Doherty, J., 2005. A hybrid regularized inversion methodology for highly parameterized environmental models. Water Resour. Res. 41 (10) https://doi.org/10.1029/2005WR003995.

Tso, C.-H., Johnson, T.C., Song, X., Chen, X., Kuras, O., Wilkinson, P., Uhlemann, S., Chambers, J., Binley, A., 2020. Integrated hydrogeophysical modelling and data assimilation for geoelectrical leak detection. J. Contam. Hydrol. 234, 103679. https://doi.org/10.1016/j.jconhyd.2020.103679.

Vilhelmsen, T.N. and Ferré, T. "Extending Data Worth Analyses to Select Multiple Observations Targeting Multiple Forecasts." Groundwater (Wiley Online Library), 2017.

Wagner, B.J., 1992. Simultaneous parameter estimation and contaminant source characterization for coupled groundwater flow and contaminant transport modelling. J. Hydrol. (Elsevier) 135 (1-4), 275–303.

White, J.T., Fienen, M.N., Doherty, J.E., 2016. A python framework for environmental model uncertainty analysis. Environ. Modell. Software 85, 217–228.

Xu, T., Gómez-Hernández, J.J., 2016. Joint identification of contaminant source location, initial release time, and initial solute concentration in an aquifer via ensemble Kalman filtering. Water Resour Res. (Wiley Online Library) 52 (8), 6587–6595.

Xu, T., Gómez-Hernández, J.J., 2018. Simultaneous identification of a contaminant source and hydraulic conductivity via the restart normal-score ensemble Kalman filter. Advances in Water Resources (Elsevier) 112, 106–123.

Xue, L., Zhang, D., Guadagnini, A., Neuman, S.P., 2014. Multimodel B ayesian analysis of groundwater data worth. Water Resour. Res. 50 (11), 8481–8496.