Combining geostatistics and simulations of flow and transport to characterize contamination within the unsaturated zone

Léa Pannecoucke, Mathieu Le Coz, Xavier Freulon, Chantal de Fouquet

Combining geostatistics and simulations of flow and transport to characterize contamination within the unsaturated zone

Léa Pannecoucke[a,*], Mathieu Le Coz[b], Xavier Freulon[a], Chantal de Fouquet[a]

[a]*MINES ParisTech, PSL University, Centre de Géosciences, 35 rue St Honoré, 77300 Fontainebleau, France*
[b]*Institut de Radioprotection et de Sûreté Nucléaire (IRSN), PSE-ENV/SEDRE, 31 avenue de la Division Leclerc, 92260 Fontenay-aux-Roses, France*

Abstract

Characterization of contamination in soils or groundwater resulting from industrial activities is critical for site remediation. In this study, geostatistics and physically-based simulations are combined for estimating levels of contamination within the unsaturated zone. First, a large number of flow and transport simulations are run and their outputs are used to compute empirical non-stationary variograms. Then, these empirical variograms, called numerical variograms and which are expected to reproduce the spatial variability of the contaminant plume better than a usual variogram model based on observations only, are used for kriging.

The method is illustrated on a two-dimensional synthetic reference test case, with a contamination due to a point source of tritium (*e.g.* tritiated water). The diversity among the simulated tritium plumes is induced by numerous sets of hydraulic parameter fields conditioned by samples from the

*Corresponding author: lea.pannecoucke@mines-paristech.fr

reference test case. Kriging with numerical variograms is then compared to ordinary kriging and kriging with an external drift: the results show that kriging with numerical variograms improves the estimates, all the more that few observations are available, underlining the interest of the method. When considering a relatively dense sampling scenario, the mean absolute error with kriging with numerical variograms is reduced by 52% compared to ordinary kriging and by 45% compared to kriging with an external drift. For a scarcer sampling, those errors are respectively reduced of 73% and 34%. However, the performance of the method regarding the classification into contaminated or not contaminated zones depends on the pollution threshold. Yet, the distribution of contamination is better reproduced by kriging with numerical variograms than by ordinary kriging or kriging with an external drift.
Keywords: Soil hydraulic parameters, Unsaturated zone, Tritium plume, Parameters uncertainties, Empirical variogram, Random fields.

---

1. Introduction

Characterization of contamination resulting from industrial activities in soils or groundwater is a major issue for site remediation (Last et al., 2004; Zhang et al., 2010). The extent and level of the potential contamination should be known as precisely as possible, with minimum uncertainty. This is an essential condition to provide appropriate decision support systems and to reduce environmental, economic and societal risks (Schädler et al., 2011; Chen et al., 2019).

Kriging is used to map contamination in soils and groundwater as it provides linear and unbiased estimates of pollutant concentration at unsampled locations (*e.g.*, Demougeot-Renard et al., 2004; Saby et al., 2006; Juang et al., 2008; D'Or et al., 2009; Pelillo et al., 2014; Liang et al., 2018). However, the quality of the kriging estimator strongly depends on its ability to model the spatial structure of the studied variable through the variogram or the covariance function. In particular, the kriging estimator is often poorly accurate if the number of sampled values is low or if the spatial variability of the studied variable is governed by complex processes (Webster and Oliver, 2007; Wang et al., 2017). Besides, the standard kriging estimator does not take into account knowledge on flow and transport processes: contamination maps obtained by kriging are not necessarily consistent with flow and transport equations.

Physically-based simulations of flow and solute transport are another widely used approach to assess contaminated soils and groundwater (*e.g.*, Neukum and Azzam, 2009; Bugai et al., 2012; Cadini et al., 2016; Testoni et al., 2017). Such simulations take into account complex processes governing contamination spread but they require a relevant definition of initial and boundary conditions, as well as internal hydraulic properties. Within the unsaturated zone, the inference of these hydraulic properties is difficult, time-consuming (Schaap et al., 2004) and the induced uncertainties result in a lack of accuracy in the characterization of the contaminated areas (Pannecoucke et al., 2019).

Various strategies have been proposed to combine kriging and physically-based simulations in order to incorporate physical behavior as expressed in flow and transport models and spatial correlation as quantified by geostatistical modeling. For example, Rivest et al. (2008) interpolate hydraulic heads using outputs from flow simulations as an external drift for constraining kriging; Shlomi and Michalak (2007) reproduce a groundwater contaminant plume by assimilating the covariance of the measured concentrations within the inversion procedure of a flow and transport model. In those studies, the geostatistical properties of the spatial variable are estimated from measurements.

Roth (1995) and Roth et al. (1998) propose to compute empirical covariances of hydraulic head within a saturated zone from a set of flow simulation outputs; Schwede and Cirpka (2010) compute the prior statistical properties of solute concentration in groundwater from Monte Carlo flow and transport steady-state simulations. The approach appears to be more suitable when the physically-based simulations do not result in a clear trend or when a large number of unknown parameters hampers the inversion of flow and transport model.

The present study aims at combining kriging and flow and transport simulations, by computing variograms from simulation outputs (called numerical variograms), in order to improve the characterization of a contaminant plume under a complex configuration, *i.e.*, by considering transient unsaturated flow and highly variable hydraulic properties. First, the geostatistical framework and the numerical variograms method are presented (section 2). Then, a two-dimensional (2D) synthetic test case is built to assess the performance of the method (section 3). The global process of implementing kriging with numerical variograms is then detailed on this test case (section 4). Finally, results are presented (section 5) and discussed (section 6).

2. Kriging with numerical variograms

This section recalls the principle of ordinary kriging estimator and introduces the numerical variograms method.

2.1. Geostatistical framework: ordinary kriging

Ordinary kriging is widely used to map pollutant concentrations in soil and groundwater. The estimate of the variable of interest Z at a target point x_0, $Z^*(x_0)$, is a linear combination of the observations:

$$Z^*(x_0) = \sum_{a=1}^{N} \lambda_a Z(x_a), \tag{1}$$

where λ_a are the kriging weights to be determined and x_a are the locations of the N observations. Ordinary kriging assumes that (i) the mean of the regionalized variable (Z) under study is constant but unknown; and (ii) the variance of any increments, *i.e.* the variogram function $\gamma(x, x') = \frac{1}{2}Var\{[Z(x) - Z(x')]^2\}$, is known for any pairs of points in the studied domain. The unbiasedness condition $E[Z(x_0) - Z^*(x_0)]$ and the minimization of the error variance $Var[Z(x_0) - Z^*(x_0)]$ define the kriging system (Chilès and Delfiner, 2012):

$$\begin{bmatrix} -\Gamma & \mathbf{1} \\ \mathbf{1}^t & 0 \end{bmatrix} \begin{bmatrix} \Lambda \\ \mu \end{bmatrix} = \begin{bmatrix} -\Gamma_0 \\ 1 \end{bmatrix}, \tag{2}$$

where $\Gamma = [\gamma(x_a, x_b)]$ is the matrix of variogram between each pair of observations (size NxN), $\mathbf{1}$ is a vector of unit values (size N), $\Lambda = [\lambda_a]$ is the vector of kriging weights, μ is a Lagrange parameter and $\Gamma_0 = [\gamma(x_a, x_0)]$

is the vector of variogram between the observations and the target point. In addition, the kriging error variance is given by:

$$\sigma_K^2(x_0) = Var[Z(x_0) - Z^*(x_0)] = \begin{bmatrix} \Lambda \\ \mu \end{bmatrix}^t \begin{bmatrix} \Gamma_0 \\ 1 \end{bmatrix}. \tag{3}$$

Hence, solving the kriging system requires the variogram values between each pair of observations and between the target and the observations. Generally, the experimental variogram is computed using the observations and then a variogram model is fitted.

However, the experimental variogram may be instable when only few data are available. In addition, the experimental variogram relies on several assumptions about the regionalized variable under study, such as stationarity or isotropy. Therefore, expert knowledge might be taken into account to improve the variogram fitting (Chilès and Delfiner, 2012).

2.2. Numerical variograms

Instead of computing the experimental variogram from observations, non stationary numerical variograms are computed from several realizations of Z. For the application presented in this study, these realizations result from a physically-based model, *e.g.*, flow and transport simulations of a contaminant plume. The numerical variogram $\hat{\gamma}$ between two points x and x' is the average of the increments computed on the realizations:

$$\hat{\gamma}(x, x') = \frac{1}{P} \sum_{p=1}^{P} \frac{1}{2} [Z_p(x) - Z_p(x')]^2, \tag{4}$$

where $Z_p(x)$ (resp. $Z_p(x')$) is the value of Z at location x (resp. x') for the p-th realization.

The object defined in Eq. (4) is a proper variogram, since it is conditionally definite-positive (Chilès and Delfiner, 2012). Indeed, it satisfies the condition:

$$-\sum_{i=1}^{M}\sum_{j=1}^{M}\omega_i\omega_j\hat{\gamma}(x_i,x_j)=\frac{1}{P}\sum_{p=1}^{P}\Big[\sum_{i=1}^{M}\omega_i Z_p(x_i)\Big]^2\geq 0 \tag{5}$$

for all $(x_i)_{i=1,\ldots,M}$, for all $(\omega_i)_{i=1,\ldots,M}$ such that $\sum_{i=1}^{M}\omega_i=0$ and for all M (de Fouquet, 2019). It ensures the consistency of the kriging system and the variogram values can then be computed for each pair of points (x,x') needed to build the matrices Γ and Γ_0. In this method, the variogram is assumed to exist and the mean of Z is assumed to be constant. The latter assumption might appear too constraining and a slightly different approach that takes into account the spatial variability of the mean of Z is presented in Appendix 1.

Numerical variograms are expected to reproduce the spatial variability of Z better than a model based on observations only, since they use physically-based simulations. More precisely, Z results from the application of a non-linear operator H on a set of inputs Y: $Z=H(Y)$. The variability among the realizations of Z is induced by the variability of Y (the randomization of the inputs Y is presented in Appendix 2). In the case of flow and transport modeling, some input parameters, such as hydraulic properties, are more difficult to determine than others. Consequently, the set of simulations should take into account the uncertainties on those parameters, by considering different input scenarios and thus covering the range of possible cases.

3. A reference test case

In this section, a synthetic reference test case is built to assess kriging with numerical variograms. This reference case consists in a two-dimensional (2D) vertical plane of 100 m large by 15 m deep in an unsaturated zone contaminated with a point source of tritiated water. The generation of this reference case is composed of three steps: (i) generation of textural properties of the surficial formation; (ii) conversion of these properties into hydraulic parameter fields; and (iii) simulation of a tritium plume with a flow and transport numerical code.

3.1. Textural properties

The surficial formation is assumed to be composed of a single facies with a spatially variable texture. The proportions of sand, silt and clay are considered to follow a normal distribution (*e.g.*, Reza et al., 2015; Usowicz and Lipiec, 2017; Taye et al., 2018) and the spatial variability in these proportions is modeled by an exponential variogram with an anisotropy between the horizontal and vertical directions (*e.g.*, Reza et al., 2015; Usowicz and Lipiec, 2017). A triplet of random fields specifying sand, silt and clay contents with a 0.5 m x 0.5 m spatial resolution is generated following the previous assumptions, using the turning bands method (Lantuéjoul, 2002). The mean ($\pm$ standard deviation) of the sand, silt and clay proportions are set to 75% ($\pm$ 10%), 12.5% ($\pm$ 6%) and 12.5% ($\pm$ 6%) and correlation lengths of 10 m and 3 m are considered in the horizontal and vertical directions respectively (Figure 1a).

3.2. Hydraulic parameters

In the unsaturated zone, flow processes are strongly related to the moisture retention curve and the relative hydraulic conductivity function. The Mualem-van Genuchten (MvG) model (Mualem, 1976; van Genuchten, 1980) describes the links between water pressure head (ψ), water content (θ) and relative hydraulic conductivity (K):

$$\theta(\psi) = \theta_r + \frac{\theta_s - \theta_r}{(1 + |\alpha\psi|^n)^m} \qquad \text{with} \qquad m = 1 - \frac{1}{n}, \tag{6}$$

and

$$K(\psi) = K_s S_e^{\frac{1}{2}}[1 - (1 - S_e^{\frac{1}{m}})^m]^2 \qquad \text{with} \qquad S_e = \frac{\theta(\psi) - \theta_r}{\theta_s - \theta_r}, \tag{7}$$

where θ_r and θ_s are respectively the residual and saturated volumetric water contents [L^3.L^{-3}], α is inversely proportional to the air-entry value [L^{-1}], n is a pore-size distribution index [-] and K_s is the saturated hydraulic conductivity tensor [L.T^{-1}].

Since the measurement of MvG parameters is complex (Schaap et al., 2004), they are commonly estimated from textural properties, which measurements are easier (e.g., Wösten et al., 1999; Tóth et al., 2015; Zhang and Schaap, 2017). The relationships linking MvG parameters and textural properties, called pedotransfer functions (PTF), are mostly based on regression analysis of existing soil databases. In this study, the random fields describing the textural properties of the surficial formation are converted into five MvG parameter fields (K_s, α, n, θ_r and θ_s) by means of rosetta3 (Zhang and Schaap, 2017). For given sand, silt and clay contents, the average values of MvG parameters are considered (Figure 1b).

3.3. Tritium plume modeling

The generated MvG parameter fields are used as input to a numerical code that simulates flow and solute transport. The tritium plume is simulated with MELODIE code, which is developed by the French Institute for Radiation Protection and Nuclear Safety (IRSN). This code simulates underground flow and solute transport in saturated and unsaturated porous media within the framework of radioactive waste disposal facilities (IRSN, 2009; Amor et al., 2014; Amor et al., 2015; Bouzid et al., 2018). MELODIE is set for solving in 2D the Richards equation describing flow in variably saturated porous media and an advection-dispersion-reaction equation representing the migration of radionuclides (Pannecoucke et al., 2019). The modeling domain is discretized in triangles with 0.5 m base and 0.25 m height. The five MvG parameter fields define the hydraulic properties within the domain. The boundary conditions are set as follows:

1. a fixed hydraulic head corresponding to the mean water table elevation (7.5 m above the bottom boundary with a 0.004 $m.m^{-1}$ lateral gradient) is set on both sides of the domain;
2. no-flow conditions are set on the bottom boundary;
3. a time variable flow corresponding to the daily percolation rate, typical from center of France, and estimated from the water balance method (Thornthwaite and Mather, 1955) is imposed on the top boundary.

A point source of tritiated water is simulated by setting an activity of 1,000 $Bq.d^{-1}$ during one month on the top surface node on the center of the modeling domain. The evolution of the activity within the domain is then simulated during five years with an adaptive time stepping (from 10^{-20} to 1 d) by considering a retardation factor of 1 and a decay constant of $1.54.10^{-4}$ d^{-1} (Figure 1c).

3.4. Reference dataset

Two types of observations are extracted from the synthetic test case, in accordance with a potential decommissioning case.

1. The texture is sampled in 8 boreholes crossing the unsaturated zone (7 m deep) distributed over the whole modeling domain (Figure 1a). Those boreholes are assumed to provide accurate observations of sand, silt and clay contents with 0.5 m vertical resolution.
2. The tritium plume is sampled to obtain observations of volumic activity with 0.5 m vertical resolution within boreholes crossing the unsaturated zone (7 m deep). Two sampling scenarios are considered: (i) 7 boreholes distributed over a zone of 20 m wide around the tritium source (scenario S1, Figure 1d); and (ii) 4 boreholes distributed over the same zone (scenario S2, Figure 1e). It is interesting to notice that for sampling scenario S2, the high values of activity are not sampled, contrary to sampling scenario S1.

Besides, an additional test case is built using the same model settings (initial and boundary conditions) but another realization of the texture fields. It results in a plume with a different shape from the reference (Figure 2).

4. Estimation by kriging with numerical variograms

In this section, kriging with numerical variograms (KNV) is carried out to estimate the tritium activity of the plume modeled in section 3, from the observations of volumic activity previously sampled. First, hydraulic parameters random fields are generated from the punctual texture observations available in the reference dataset (section 3.4). Then, 2,000 unconditioned tritium plumes are simulated by means of a flow and transport model. These simulations are used to compute numerical variograms of activity and finally interpolate punctual activity observations from the reference dataset (according to scenarios S1 or S2, section 3.4).

4.1. Hydraulic parameters random fields

A large number of random fields describing the MvG hydraulic parameters (K_s, α, n, θ_r, θ_s) within the surficial formation are generated based on two different approaches.

1. Approach 1: the observations of sand, silt and clay contents available in the reference dataset are used to compute experimental variograms, which allow the generation of 1,000 triplets of conditional fields of sand, silt and clay contents. The variogram parameters are randomized (see Appendix 2) and the conditional simulations follow the distribution (close to normal) given by the observations from the reference dataset. The resulting triplets of random fields describing the textural properties are converted into MvG parameter fields using rosetta3 PTF (for given sand, silt and clay contents, the average values of MvG parameters are considered). It results in 1,000 sets of 5 random fields.
2. Approach 2: the sand, silt and clay contents available in the reference dataset are converted into MvG parameters using rosetta3 PTF (for given sand, silt and clay contents, the average values of MvG parameters are considered). Experimental variograms are computed from

these values of MvG parameters, which are then interpolated by means of a conditional simulation tool considering variogram models with randomized parameters (Appendix 2). Normal distributions of θ_r and θ_s and lognormal distributions of K_s, α and n are considered (*e.g.*, Botros et al., 2009; Pannecoucke et al., 2019), with means and variances given by the values of hydraulic parameters at sampled locations. It results in 1,000 sets of 5 random fields.

4.2. Simulations of flow and solute transport

The MvG parameter fields obtained via the two previous approaches are set as inputs to MELODIE code to simulate 2,000 tritium plumes. All the other model parameters are kept constant compared to the test case described in section 3.3.

4.3. Estimation and performance assessment

The set of 2,000 simulated plumes is used to compute the numerical variograms between each couple of points needed to build the kriging system. The KNV estimate is computed using (i) the observations from 7 boreholes (S1); and (ii) the observations from 4 boreholes (S2).

Two other kriging methods are used as benchmarks: (i) ordinary kriging (OK), with a stationary model of variogram based on the observations only; and (ii) kriging with an external drift (KED) with auxiliary variables given by simulation outputs (Rivest et al., 2008). More precisely, the empirical average of the simulations (mean plume) is used as an auxiliary variable, and thus the empirical mean of Z is considered variable over the modeling domain (see Appendix 1).

In order to assess the performances of KNV compared to OK and KED, several indicators are computed.

1. The maps of estimation, estimation error and kriging error standard deviation are computed. For OK and KED, the maps of kriging error standard deviation are corrected by a proportional effect (Donati and de Fouquet, 2018) in order to account for the zones of low or high values of estimated activity. This supplementary modeling step is not needed for KNV, because numerical variograms directly account for the local variability of activity in the contaminated zone.
2. The errors are quantified in terms of mean absolute error (MAE), root-mean-square error (RMSE) and mean relative error (MRE). The MRE is given by:

$$MRE = \frac{1}{n_{cells}} \sum_{i=1}^{n_{cells}} \frac{Z^{ref}(x_i) - Z^*(x_i)}{max(1, Z^{ref}(x_i))} \tag{8}$$

   where n_{cells} is the number of cells in the modeling domain (without the observations), $Z^{ref}(x_i)$ (resp. $Z^*(x_i)$) is the value of activity of the reference plume (resp. the estimation) at location x_i. The denominator is set to 1 if $Z^{ref}(x_i) \leq 1$ in order to avoid huge relative errors when $Z^{ref}(x_i)$ is close to 0.
3. The ability of the estimator to reproduce the distribution of the actual contamination is assessed via the selectivity curve (Chilès and Delfiner, 2012). This curve is parametrized by the contamination threshold z. For each z, two quantities are computed.
   - the percentage of grid cells in the modeling domain such that $Z(x_i) \geq z$ (on the x-axis), defined as:

$$\frac{\sum_{i=1}^{n_{cells}} \mathbb{1}_{Z(x_i)\geq z}}{n_{cells}} \times 100 \tag{9}$$

where $\mathbb{1}_{Z(x_i)\geq z}$ equals 1 if $Z(x_i) \geq z$, 0 otherwise;

- the corresponding percentage of total volumic activity contained by the previous grid cells (on the y-axis), defined as:

$$\frac{\sum_{i=1}^{n_{cells}} Z(x_i)\mathbb{1}_{Z(x_i)\geq z}}{\sum_{i=1}^{n_{cells}} Z(x_i)} \times 100. \tag{10}$$

4. The proportions of false-positive and false-negative surfaces are computed for several contamination thresholds (z). The proportion of false-positive surface is defined as the number of grid cells such that $Z^*(x_i) \geq z$ and $Z^{ref}(x_i) < z$, divided by the number of grid cells such that $Z^{ref}(x_i) \geq z$ (the actual surface of the contaminated zone on the reference, which depends on the contamination threshold). The proportion of false-negative surface is defined as the number of grid cells such that $Z^*(x_i) < z$ and $Z^{ref}(x_i) \geq z$, divided by the actual surface of the contaminated zone. This indicator assesses the risk of leaving on-site contamination (false-negative) or on the contrary of overestimating the extent of the contamination and the associated remediation costs (false-positive).

5. Results

In this section, the performance indicators described above are computed for the estimates of the reference plume obtained by OK, KED, KNV and for sampling scenarios S1 and S2. Then, the results are presented for the additional test case (section 3.4, Figure 2). Finally, the KNV estimates computed when distinguishing the two sets of simulations based on approach 1 or approach 2 (section 4.1) are compared.

5.1. Sampling scenario S1

The maps of estimation are almost similar (Figures 3a, 3b, 3c) for the three methods. Yet, the errors are slightly higher for OK and KED than for KNV (Figures 3d, 3e, 3f). Besides, the theoretical standard deviations of kriging error are much higher for OK and KED than for KNV, even when a proportional effect is taken into account (Figures 3g, 3h, 3i). In accordance with this qualitative assessment, KNV results in smaller mean errors than OK and in a lesser extent KED (Table 1), whatever the actual activity values (Figure 4a).

The selectivity curves show that KNV estimate slightly better reproduces the actual distribution of activity than OK and KED estimates (Figure 4b). The curves obtained with the three approaches are yet almost overlaying each other.

The proportion of false-positive surface is smaller for KNV that for OK, whatever the contamination threshold (Figure 4c). This proportion is reduced of 10%, except for contamination thresholds above 1,000 Bq.m$^{-3}{}_{H2O}$ (mainly because the contaminated surfaces are more and more reduced when the threshold increases). The proportion of false-positive surface is yet smaller for KED than for KNV for very low contamination thresholds (below 20 Bq.m$^{-3}{}_{H2O}$); for higher contamination thresholds, KNV leads to smaller proportion of false-positive surfaces than KED. The proportion of false-negative surface is slightly higher for KNV than for OK and KED for contamination thresholds below 500 $Bq.m^{-3}{}_{H2O}$ (Figure 4d). For higher contamination thresholds, KNV performs better than OK and KED, because numerical variograms are non stationary and enables a better estimation of high values of activity.

5.2. Sampling scenario S2

For sampling scenario S2, the maps of estimation obtained by the three approaches look different (Figures 5a, 5b, 5c). The shape of the plume estimated by OK appears poorly consistent, while the plumes estimated by KED and KNV respect the global shape of the reference plume. Yet, the plume estimated by KED is more attenuated than the one obtained by KNV. Besides, standard deviations of kriging error are higher for OK and even more for KED than for KNV (Figures 5g, 5h, 5i). MAE, RMSE and MRE are smaller for KNV than for OK and KED (Table 1). In particular, OK and KED tend to reduce the actual variability of activities (overestimation of lowest activities, underestimation of highest activities), while KNV results in a more consistent distribution of activities, despite an overall overestimation, especially for the highest values of activity (Figure 6a).

The selectivity curves show that KNV and KED better reproduce the actual distribution of activity than OK (Figure 6b). For example, 10% of the modeling surface contains 80% of the whole contamination for KED and KNV estimates, while 18% of the modeling domain contains the same amount of contamination for the activity field estimated by OK.

The false-positive surfaces obtained by KNV are smaller than the ones obtained by OK and KED (Figure 6c), except for contamination thresholds higher than 1,000 Bq.m$^{-3}{}_{H2O}$ (due to the fact that OK and KED tend to underestimate high values of activity while KNV overestimates high values of activity). The false-negative surfaces obtained by KNV are generally larger than the ones obtained by OK and KED, at least for contamination thresholds below 800 Bq.m$^{-3}{}_{H2O}$ (Figure 6d).

5.3. Additional test case

In order to test the reproducibility of the proposed approach, the same study has been made on the additional test plume (Figure 2), which has a more complex shape than the reference one. MAE, RMSE and MRE are reduced for KNV, compared to OK and KED (Table 2, Figures 7a and 8a) for sampling scenarios S1 and S2. Contrary to the reference test case, the errors are higher for KED than for OK (Table 2).

The selectivity curves (Figure 7b and 8b) show that KNV better reproduces the distribution of the actual contamination than KED and OK, especially for sampling scenario S2.

The false-positive surface is smaller for KNV than for OK and KED, for both sampling scenarios (Figure 7c and 8c). For the false-negative surfaces, the performances of each method depend on the contamination threshold. For S1, for low thresholds (below 50 Bq.m$^{-3}{}_{H2O}$) KED performs better than OK and than KNV, while for higher thresholds, KNV performs better than OK and than KED (Figure 7d). For S2, OK and KNV perform better than KED (Figure 8d).

5.4. Hydraulic parameter fields

In section 4.1, two slightly different approaches have been introduced to generate MvG parameter random fields. For the results presented above, the simulated plumes obtained via the two approaches have been gathered and mixed to compute numerical variograms. In order to compare both approaches, KNV is implemented with (i) numerical variograms computed from 1,000 simulations generated with approach 1 (KNV-1); (ii) numerical variograms computed from 1,000 simulations generated with approach 2 (KNV-2).

For the reference test case, the estimated plumes are almost the same for KNV-1, KNV-2 and KNV. Indeed, MAE are really close, especially for S1 (Table 3). For S2, KNV-2 leads to smaller errors than KNV and than KNV-1. For the additional test case, the results obtained with KNV-2 and KNV are almost similar (Table 3). On the contrary, the results obtained with KNV-1 are unacceptable (the estimated plumes are not consistent at all), for both sampling scenarios.

6. Discussion

Spatial variability of MvG parameters is generally poorly characterized at field scale even if it can significantly affect the evolution of contaminant plumes within the unsaturated zone (Pannecoucke et al., 2019). For example, in this study, the tritium plumes simulated using a similar groundwater flow and transport model but various MvG parameter fields (generated from observations of texture sampled in 8 boreholes) are significantly different: their surfaces range from 60 to 150 m^2 and their mass centers are spread over 20 m wide (Figure 9). Therefore, although the initial and boundary conditions of the flow and transport model are fixed, the uncertainties related to hydraulic parameters within the surficial formation do not lead to an accurate characterization of the contamination.

To improve plume characterization and delineation, kriging with numerical variograms, consisting in using flow and transport simulation outputs to compute numerical variograms, appears to perform better than standard geostatistical tools (OK and KED), at least for most of the various indicators considered in this study. KNV appears to be particularly interesting when the available observations are scarce, as shown by the larger difference of performances between OK and KNV (or KED and KNV) for scarce (S2, 4 boreholes) compared to dense (S1, 7 boreholes) sampling scenarios. Besides, it is interesting to notice that KNV enables the estimation of high values of activity, even if those high values are not sampled, which is not the case for OK and KED (*e.g.* reference test case, scenario S2). When the actual plume differs from the mean simulated plume, KED is poorly efficient, *e.g.*, in the case of the additional test case with a more complex plume geometry.

However, for reproducing such a complex plume shape, KNV estimator results in better performances when the MvG parameter fields are generated from interpolation of punctual values of these parameters (approach 2) than from conversion of soil texture fields (approach 1). This could be explained by the fact that the approach 2 leads to a higher variability in MvG parameters and thus in more variable simulated plumes as outputs of the flow and transport model (Figure 9). A relevant characterization of the variability in hydraulic parameters therefore remains a key issue for taking advantage of

KNV. This requires to develop *in situ* approaches for better estimating soil hydraulic parameters and their variability at field scale (*e.g.*, Léger et al., 2014 and 2016).

This work focuses on uncertainties in spatial variability in MvG parameter fields. However, other input parameters, such as the location of the source of pollution or the boundary conditions, also impact flow and solute transport in the unsaturated zone. In a real study case, those parameters are not perfectly known and it would be interesting to take into account the uncertainties in those inputs.

Besides, in the case of a real contaminated site with a regulatory threshold to be respected, the delimitation into contaminated and uncontaminated zone should take into account uncertainties on the estimates, expressed by the standard deviation of kriging error, and some probabilities of exceeding a given threshold. Geostatistical conditional simulations could also have been implemented, but it requires strongest assumptions and more computational time. That is why the application was limited to estimation (as in Saby et al., 2006 or Liang et al., 2018).

7. Conclusion

This study shows that kriging with numerical variograms improves the estimates of tritium activities in the unsaturated zone. Although the assumptions might appear simplistic (stationary mean), this method leads to a reduction of the estimation errors, and more importantly of the corresponding error standard deviation (*i.e.*, more trustworthy estimators). This method is all the more interesting that the number of observations of pollutant concentration is reduced. However, the assessment procedure detailed in this study is based on a synthetic case study with well constrained boundary conditions. The next step is be to carry out the method on a real contaminated site.

In addition, the kriging with numerical variograms method can be transposed to other scales of heterogeneities, such as systems with several geological units, or other pollutants with a more complex chemical behavior, as soon as a numerical code that simulates the studied phenomenon is available. It could also be applied in completely different domains, such as air quality characterization, estimations of ocean temperatures, or population dynamics in ecology.

Acknowledgments

This study is part of *Kri-Terres* project, supported by the French National Radioactive Waste Management Agency (ANDRA) under the French "Investments for the Future" Program. The authors thank Jean-Charles Robinet (ANDRA), Clémence Houzé (UPSud GEOPS) and two anonymous reviewers for their helpful comments.

References

Amor, H., Bourgeois, M., Le, M.-H., 2014. Development of an Adaptive Mesh Refinement strategy for the Melodie software simulating flow and radionuclide transport in porous media. In: NM2 PorousMedia, October 3, 2014. Dubrovnik, Croatia.

Amor, H., Benkhaldoun, F., Bourgeois, M., Le, M.-H., 2015. Development of an adaptive mesh refinement strategy for the MELODIE software simulating flow and radionuclides transport in heterogeneous porous media. In: MAMERN VI2015 Conference, June 1-5, 2015. Pau, France.

Botros, F. E., Harter, T., Onsoy, Y. S., Tuli, A., Hopmans, J. W., 2009. Spatial Variability of Hydraulic Properties and Sediment Characteristics in a Deep Alluvial Unsaturated Zone. *Vadose Zone Journal* 8 (2), 276–289.

Bouzid, M., Espivent, C., Bourgeois, M., Amor, H., Marsal, F., 2018. Numerical modelling of radionuclide migration in the context of a near-surface LILW disposal facility. In: WM2018 Conference, March 18-22, 2018. Phoenix, Arizona, USA.

Bugai, D., Skalskyy, A., Dzhepo, S., Kubko, Y., Kashparov, V., Van Meir, N., Stammose, D., Simonucci, C., Martin-Garin, A., 2012. Radionuclide migration at experimental polygon at Red Forest waste site in Chernobyl zone. Part 2: Hydrogeological characterization and groundwater transport modeling. *Applied Geochemistry* 27 (7), 1359–1374.

Cadini, F., Tosoni, E., Zio, E., 2016. Modeling the release and transport of 90Sr radionuclides from a superficial nuclear storage facility. *Stochastic Environmental Research and Risk Assessment* 30 (2), 693–712.

Chen, I.-C., Chuo, Y.-Y., Ma, H.-W., 2019. Uncertainty analysis of remediation cost and damaged land value for brownfield investment. *Chemosphere* 220, 371–380.

Chilès, J.-P., Delfiner, P. *Geostatistics: Modeling Spatial Uncertainty, 2nd Edition*. Wiley series in probability and statistics. Wiley edition, 2012.

Demougeot-Renard, H., de Fouquet, C., Renard, P., 2004. Forecasting the Number of Soil Samples Required to Reduce Remediation Cost Uncertainty. *Journal of Environment Quality* 33 (5).

Donati, M., de Fouquet, C., 2018. How to describe organic contamination in soils: A model comparison for chlorinated solvent concentrations at industrial site scale. *Science of The Total Environment* 633, 1480–1495.

D'Or, D., Demougeot-Renard, H., Garcia, M., 2009. An Integrated Geostatistical Approach for Contaminated Site and Soil Characterisation. *Mathematical Geosciences* 41 (3), 307–322.

de Fouquet, C. *Exercices corrigés de géostatistique*. Presse des Mines, Paris, 2019.

IRSN, 2009. MELODIE Modèle d'Evaluation à LOng terme des Déchets Irradiants Enterrés. Notice théorique du code de calcul MELO version 5.0. (in french).

Juang, K.-W., Liao, W.-J., Liu, T.-L., Tsui, L., Lee, D.-Y., 2008. Additional sampling based on regulation threshold and kriging variance to reduce the probability of false delineation in a contaminated site. *Science of The Total Environment* 389 (1), 20–28.

Lantuéjoul, C. *Geostatistical Simulation - Models and Algorithms*. Springer, 2002.

Last, G. V., Rohay, V. J., Schelling, F. J., Bunn, A. L., Delamare, M. A., Dirkes, R. L., Hildebr, R. D., Morse, J. G., Napier, B. A., Riley, R. G., Soler, L., Thorne, P. D., 2004. A comprehensive and systematic approach to developing and documenting conceptual models of contaminant release and migration at the Hanford Site. *Stochastic Environmental Research and Risk Assessment* 18 (2), 109–116.

Léger, E., Saintenoy, A., Coquet, Y., 2014. Hydrodynamic parameters of a sandy soil determined by ground-penetrating radar inside a single ring infiltrometer. *Water Resources Research* 50 (7), 5459–5474.

Léger, E., Saintenoy, A., Tucholka, P., Coquet, Y., 2016. Hydrodynamic Parameters of a Sandy Soil Determined by Ground-Penetrating Radar Monitoring of Porchet Infiltrations. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9 (1), 188–200.

Liang, C.-P., Chen, J.-S., Chien, Y.-C., Chen, C.-F., 2018. Spatial analysis of the risk to human health from exposure to arsenic contaminated groundwater: A kriging approach. *Science of The Total Environment* 627, 1048–1057.

Mualem, Y., 1976. A new model for predicting the hydraulic conductivity of unsaturated porous media. *Water Resources Research* 12 (3), 513–522.

Neukum, C., Azzam, R., 2009. Quantitative assessment of intrinsic groundwater vulnerability to contamination using numerical simulations. *Science of The Total Environment* 408 (2), 245–254.

Pannecoucke, L., Le Coz, M., Houzé, C. Saintenoy, A., Cazala, C., de Fouquet, C., 2019. Impact of spatial variability in hydraulic parameters on plume migration within unsaturated surficial formations. *Journal of Hydrology* 574, 160–168.

Pelillo, V., Piper, L., Lay-Ekuakille, A., Lanzolla, A., Andria, G., Morello, R., 2014. Geostatistical approach for validating contaminated soil measurements. *Measurement* 47, 1016–1023.

Reza, S. K., Nayak, D. C., Chattopadhyay, T., Mukhopadhyay, S., Singh, S. K., Srinivasan, R., 2015. Spatial distribution of soil physical properties of alluvial soils: a geostatistical approach. *Archives of Agronomy and Soil Science* 62 (7), 972–981.

Rivest, M., Marcotte, D., Pasquier, P., 2008. Hydraulic head field estimation using kriging with an external drift: A way to consider conceptual model information *Journal of Hydrology* 361 (3), 349–361.

Roth, C. *Contribution de la géostatistique à la résolution du problème inverse en hydrogéologie*. Thèse, Ecole Nationale Supérieure des Mines de Paris, 1995.

Roth, C., Chilès, J.-P., de Fouquet, C., 1998. Combining geostatistics and flow simulators to identify transmissivity. *Advances in Water Resources* 21 (7), 555–565.

Saby, N., Arrouays, D., Boulonne, L., Jolivet, C., Pochot, A., 2006. Geostatistical assessment of Pb in soil around Paris, France. *Science of The Total Environment* 367 (1), 212–221.

Schaap, M. G., Nemes, A., van Genuchten, M. T., 2004. Comparison of

Models for Indirect Estimation of Water Retention and Available Water in Surface Soils. *Vadose Zone Journal* 3 (4), 1455–1463.

Schädler, S., Morio, M., Bartke, S., Rohr-Zänker, R., Finkel, M., 2011. Designing sustainable and economically attractive brownfield revitalization options using an integrated assessment model. *Journal of Environmental Management* 92 (3), 827–837.

Schwede, R., Cirpka, O. A., 2010. Interpolation of Steady-State Concentration Data by Inverse Modeling. *Groundwater* 48 (4), 569–579.

Shlomi, S., Michalak, A. M., 2007. A geostatistical framework for incorporating transport information in estimating the distribution of a groundwater contaminant plume. *Water Resources Research* 43 (3), W03412.

Taye, M., Simane, B., Selsssie, Y. G., Zaitchik, B., Setegn, S., 2018. Analysis of the Spatial Variability of Soil Texture in a Tropical Highland: The Case of the Jema Watershed, Northwestern Highlands of Ethiopia. *International Journal of Environmental Research and Public Health* 15 (9).

Testoni, R., Levizzari, R., De Salve, M., 2017. Coupling of unsaturated zone and saturated zone in radionuclide transport simulations. *Progress in Nuclear Energy* 95, 84–95.

Thornthwaite, C. W., Mather, J. R., 1955. The water balance. *Publications in climatology* 8, 1–104.

Tóth, B., Weynants, M., Nemes, A., Makó, A., Bilas, G., Tóth, G., 2015. New generation of hydraulic pedotransfer functions for Europe. *European Journal of Soil Science* 66 (1), 226–238.

Usowicz, B., Lipiec, J., 2017. Spatial variability of soil properties and cereal yield in a cultivated field on sandy soil. *Soil and Tillage Research* 174, 241–250.

van Genuchten, M. T., 1980. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Science Society of America Journal* 44 (5), 892–898.

Wang, Y., Akeju, O. V., Zhao, T., 2017. Interpolation of spatially varying but sparsely measured geo-data: A comparative study. *Engineering Geology* 231, 200–217.

Webster, R., Oliver, M. A. *Geostatistics for Environmental Scientists, 2nd Edition*. Statistics in Practice. Wiley edition, 2007.

Wösten, J. H. M., Lilly, A., Nemes, A., Le Bas, C., 1999. Development and use of a database of hydraulic properties of European soils. *Geoderma* 90 (3), 169–185.

Zhang, D., Shi, L., Chang, H., Yang, J., 2010. A comparative study of numerical approaches to risk assessment of contaminant transport. *Stochastic Environmental Research and Risk Assessment* 24 (7), 971–984.

Zhang, Y., Schaap, M. G., 2017. Weighted recalibration of the Rosetta pedotransfer model with improved estimates of hydraulic parameter distributions and summary statistics (Rosetta3). *Journal of Hydrology* 547, 39–53.

Appendix 1: A varying mean for Z

In section 2.2, the mean of Z is assumed to be constant. This assumption may not be consistent with the mean plume computed as the average of the simulations and used as an external drift in this study:

$$E[Z(x)] = m(x) = \frac{1}{P}\sum_{p=1}^{P} Z_p(x). \tag{.1}$$

To take into account this computed drift, a slightly different method is examined. In KNV as presented in section 2.2, a constant mean for Z leads to the following unbiasedness condition:

$$\sum_{a=1}^{N} \lambda_a = 1. \tag{.2}$$

If the mean of Z depends on the location x in the modeling domain, the unbiasedness conditions becomes:

$$\sum_{a=1}^{N} \lambda_a m(x_a) = m(x_0). \tag{.3}$$

The variance of the kriging error is given by:

$$Var[Z^*(x_0) - Z(x_0)] = \sum_{a=1}^{N}\sum_{b=1}^{N} \lambda_a \lambda_b C(x_a, x_b) - 2\sum_{a=1}^{N} \lambda_a C(x_a, x_0) + C(x_0, x_0), \tag{.4}$$

where $C(x, x')$ is the numerical covariance between x and x':

$$C(x, x') = \frac{1}{P}\sum_{p=1}^{P} [Z_p(x) - m(x)][Z_p(x') - m(x')]. \tag{.5}$$

Hence the kriging system:

$$\begin{bmatrix} C & M \\ M^t & 0 \end{bmatrix} \begin{bmatrix} \Lambda \\ \mu \end{bmatrix} = \begin{bmatrix} C_0 \\ m_0 \end{bmatrix}, \tag{.6}$$

where $C = [C(x_a, x_b)]$ is the matrix of covariances between each couple of observations, $M = [m(x_a)]$ is the vector containing the empirical means of Z at observation locations, $C_0 = [C(x_a, x_0)]$ is the vector of covariance between the target and the observations and $m_0 = m(x_0)$ is the mean of Z at the target point.

The estimates of the reference plume and the additional plume have be computed with this slightly different approach. The results are almost identical to those obtained when considering that the mean of Z is constant over the modeling domain (Figure 10). This method, which lowers the assumption of the stationary of Z, is more complex to implement than the one described in section 2.2 and does not seem to perform better.

Appendix 2: Uncertainties in the input parameters

The modeling of the uncertainties in the input parameters to the numerical code (these parameters are denoted Y in section 2.2) focuses on the hydraulic parameters fields. Thus, those fields are randomized, while the rest of the input parameters is kept constant for all simulations. To take into account the uncertainties in the experimental variograms computed from observations of sand, silt and clay contents (for approach 1) or from MvG parameters (for approach 2), the parameters of the variogram model are randomized. For each realization, the parameters of the variogram model used to simulate the fields are drawn from the following probability distributions:

1、the sill is sampled from a gaussian distribution centered on the sill of the experimental variogram with a $\pm$ 20% range of variation;

2、the vertical range is sampled from a gaussian distribution centered on the vertical range of the experimental variogram with a $\pm$ 20% range of variation;

3、the horizontal range is sampled from a triangular distribution with a mode equals to the horizontal range of the experimental variogram and the minimum and maximum values respectively to twice the vertical range and ten times the vertical range. It leads to a stronger dispersion than for the vertical range, since the inference of the horizontal range is less accurate than the vertical range due to the sampling scheme;

4、the behavior of the variogram at short distances is randomly chosen between 3 cases: a cubic model without nugget effect, an exponential model without nugget effect or an exponential model with a nugget effect (between 0 and 5% of total sill).

Table 1

Table 1: MAE [$Bq.m^{-3}_{H2O}$], RMSE [$Bq.m^{-3}_{H2O}$] and MRE [-] for both sampling scenarios and for the reference test case. Scenario S1 corresponds to 918 unknown grid cells (119 observations) and scenario S2 corresponds to 969 unknown grid cells (68 observations).

	S1			S2		
	OK	KED	KNV	OK	KED	KNV
MAE	61	53	29	173	71	47
RMSE	161	138	89	348	174	147
MRE	-4.6	-2.8	-2.2	-47	-6.8	-0.8

Table 2

Table 2: MAE [Bq.m$^{-3}{}_{H2O}$], RMSE [Bq.m$^{-3}{}_{H2O}$] and MRE [-] for both sampling scenarios and for the additional test case.

	S1			S2		
	OK	KED	KNV	OK	KED	KNV
MAE	72	119	43	139	140	82
RMSE	184	230	125	302	355	233
MRE	-4.8	-31	-2.2	-5.6	-4.2	-2.7

Table 3

Click here to download Table: Table 3.doc

Table 3: MAE [$Bq.m^{-3}_{H2O}$] for both sampling scenarios and both test cases, by differentiating KNV-1 and KNV-2 from KNV.

	Reference test case		Additional test case	
	S1	S2	S1	S2
KNV-1	30	58	163	484
KNV-2	32	41	44	92
KNV	29	47	43	82

Figure 1

Click here to download Figure: Figure1.pdf

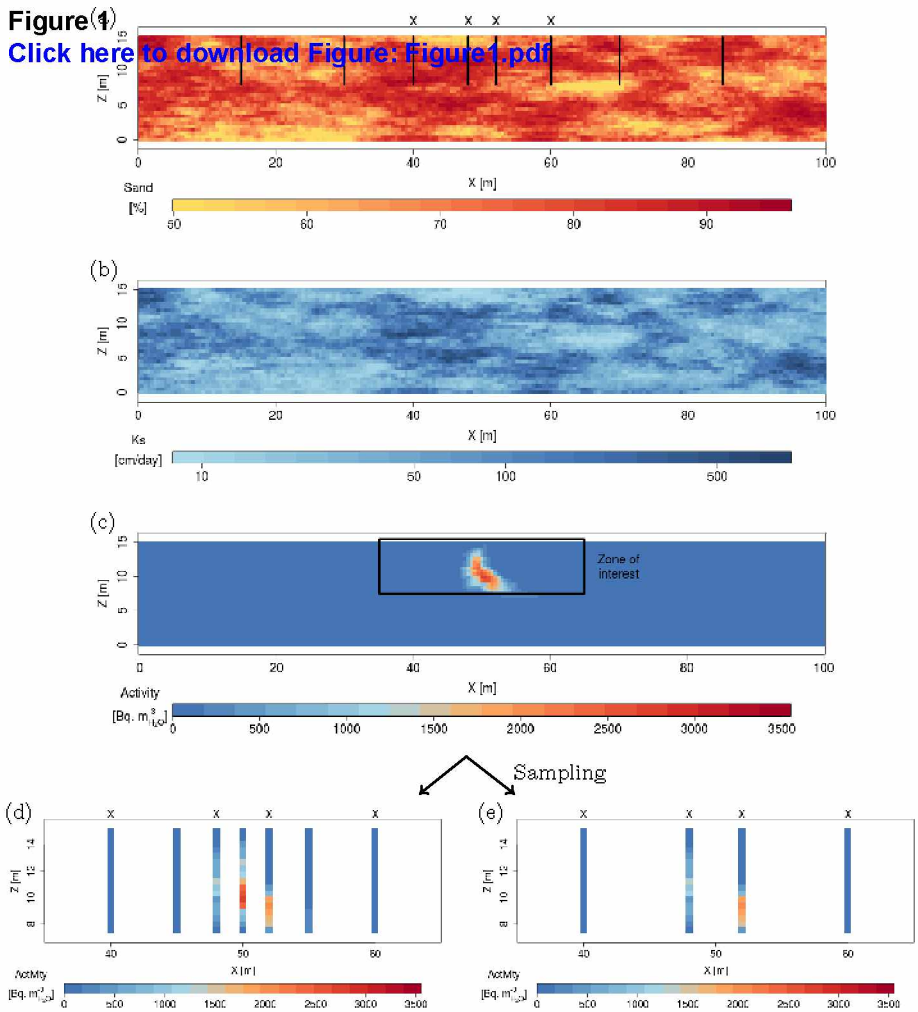

Figure 2

Click here to download Figure: Figure2.pdf

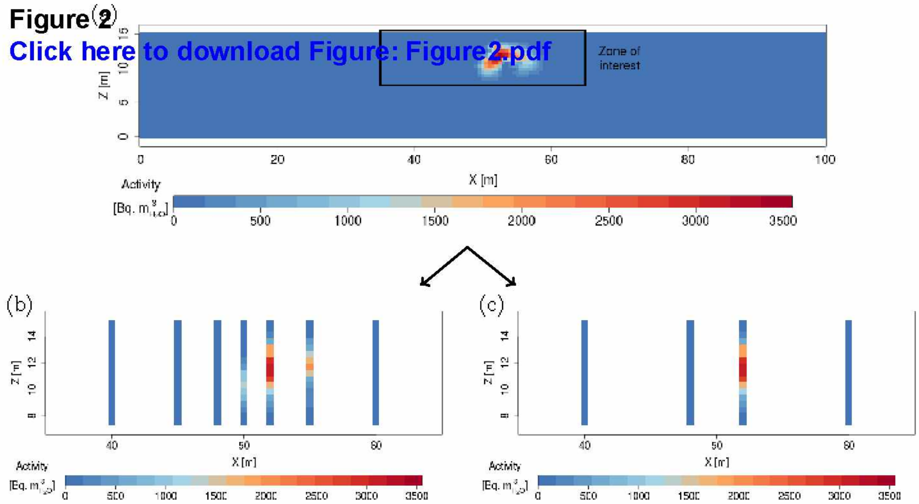

Figure 3

Click here to download Figure: Figure3.pdf

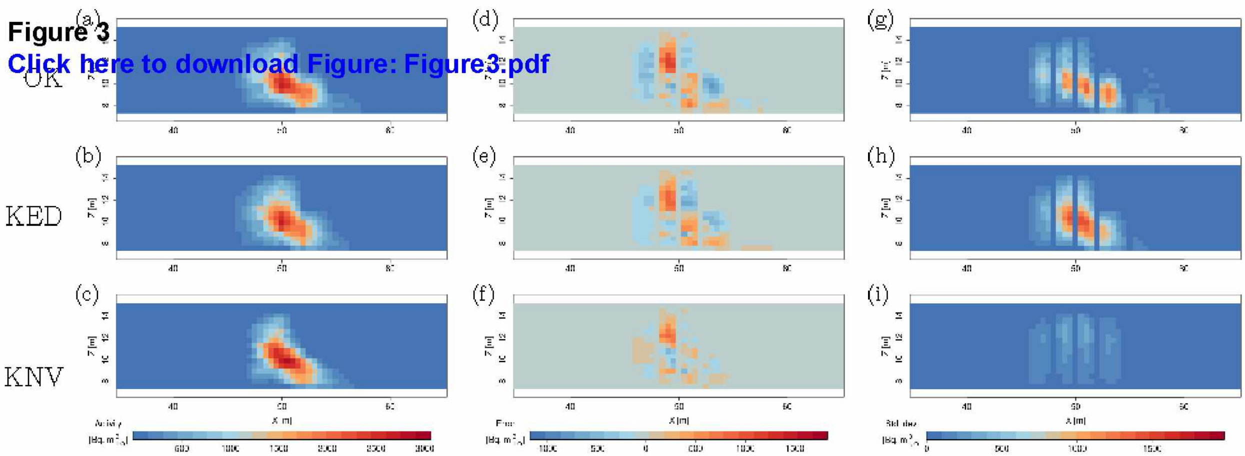

Figure 4

Click here to download Figure: Figure4.pdf

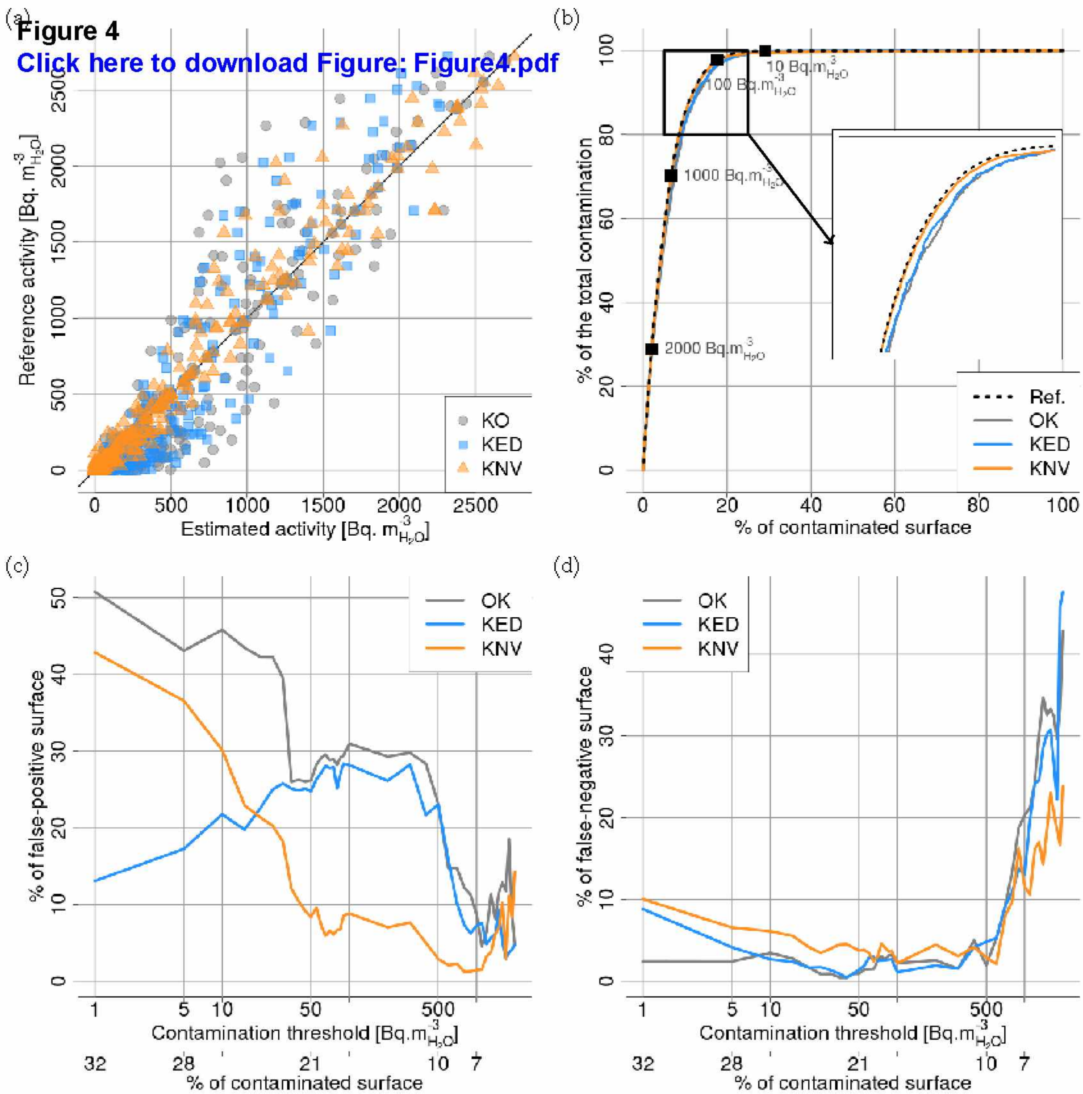

Figure 5

Click here to download Figure: Figure5.pdf

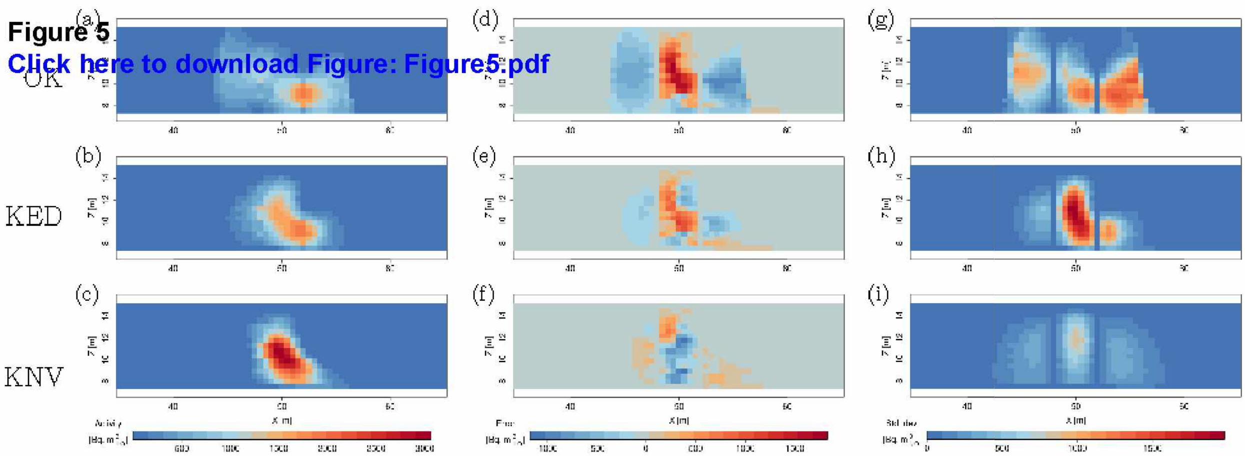

Figure 6

Click here to download Figure: Figure6.pdf

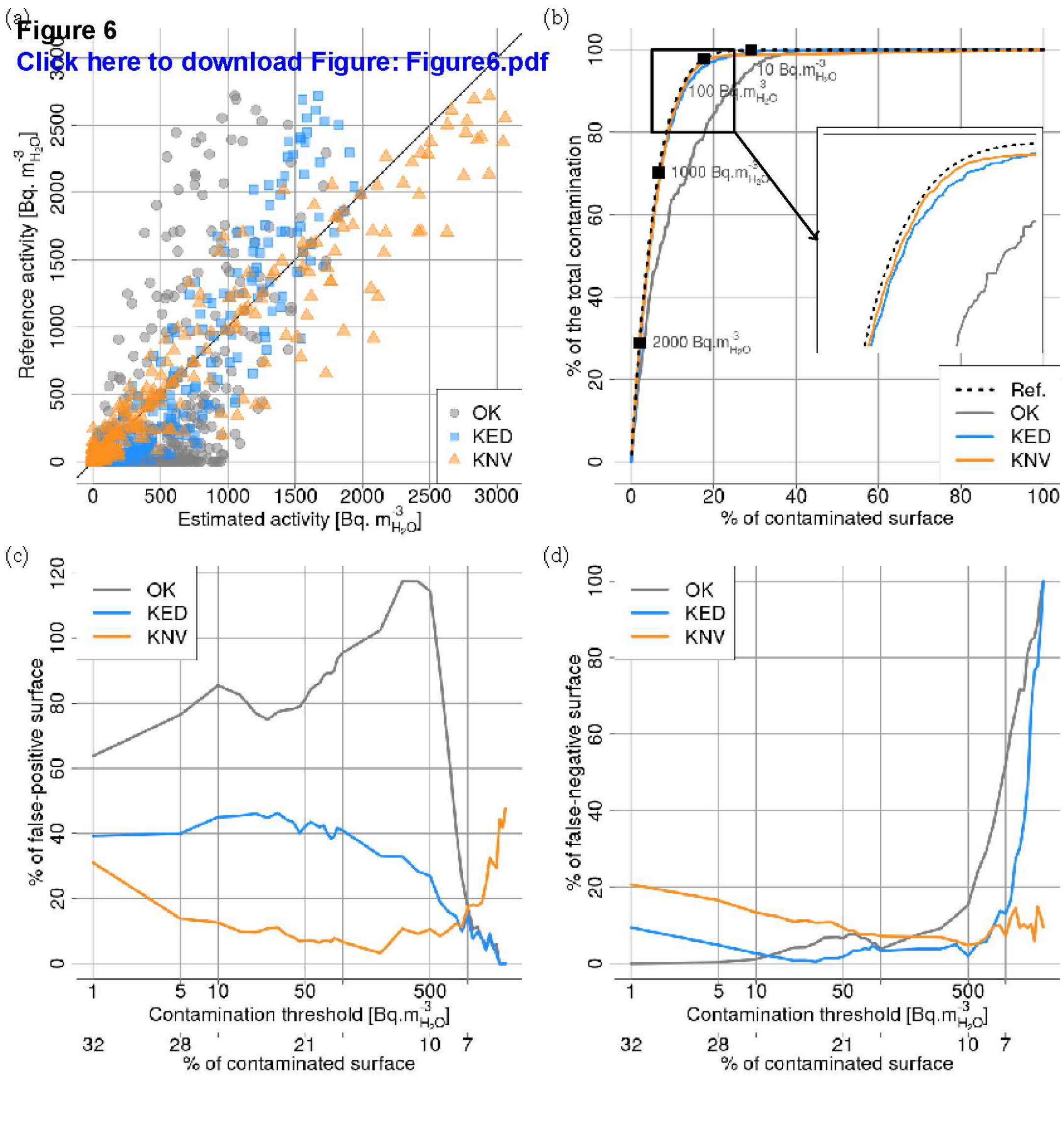

Figure 7

Click here to download Figure: Figure7.pdf

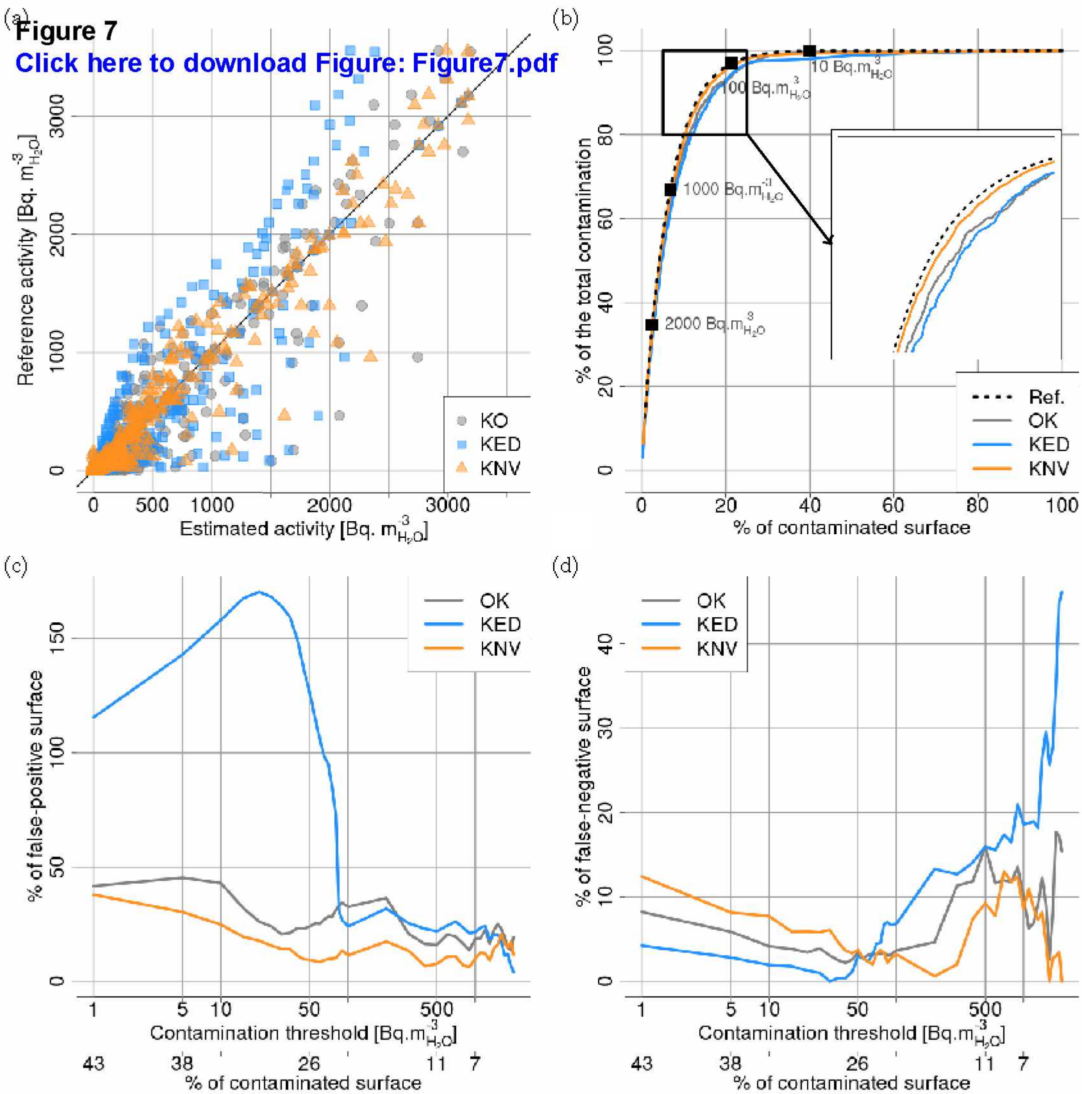

Figure 8

Click here to download Figure: Figure8.pdf

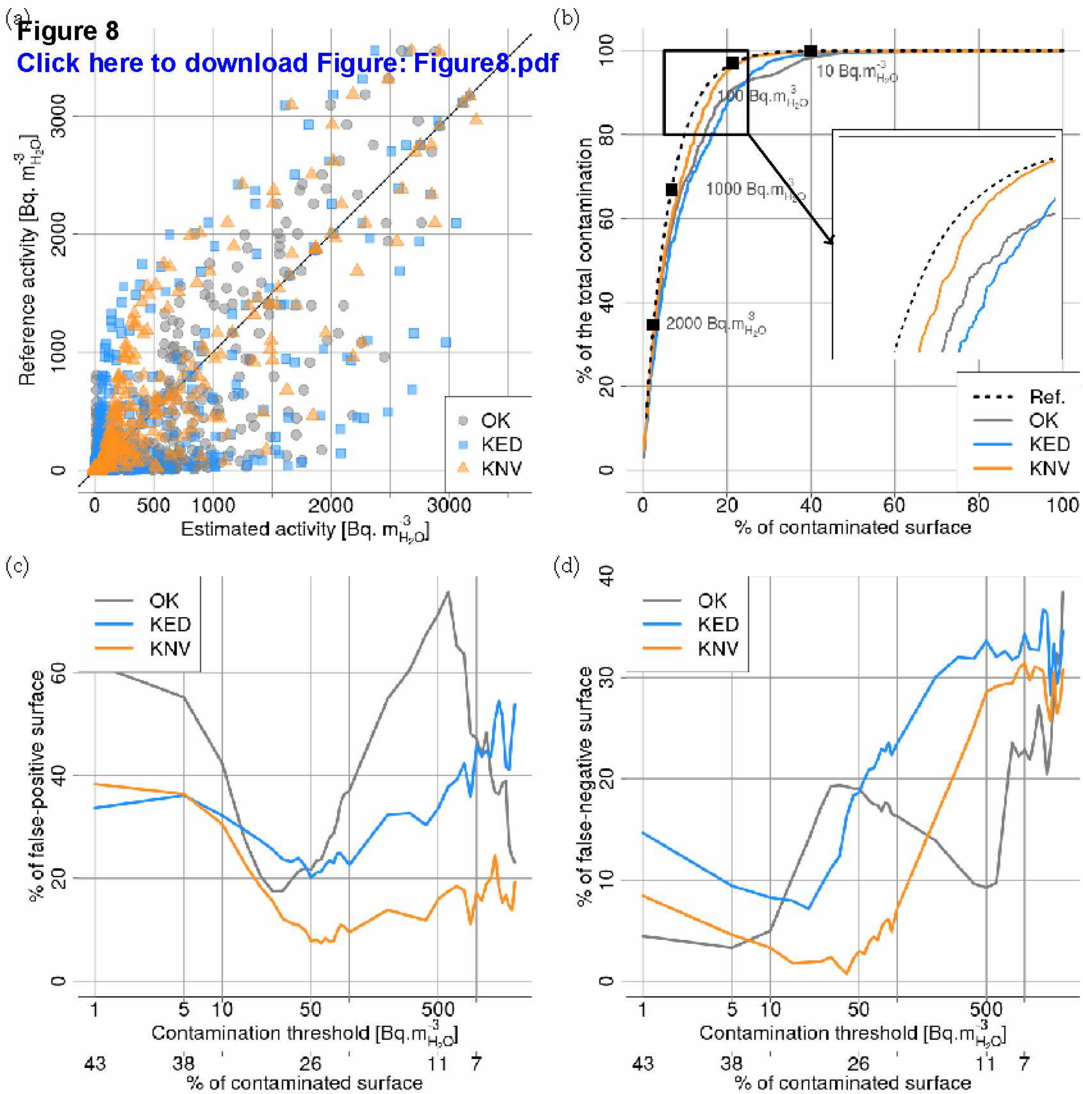

Figure 9

Click here to download Figure: Figure9.pdf

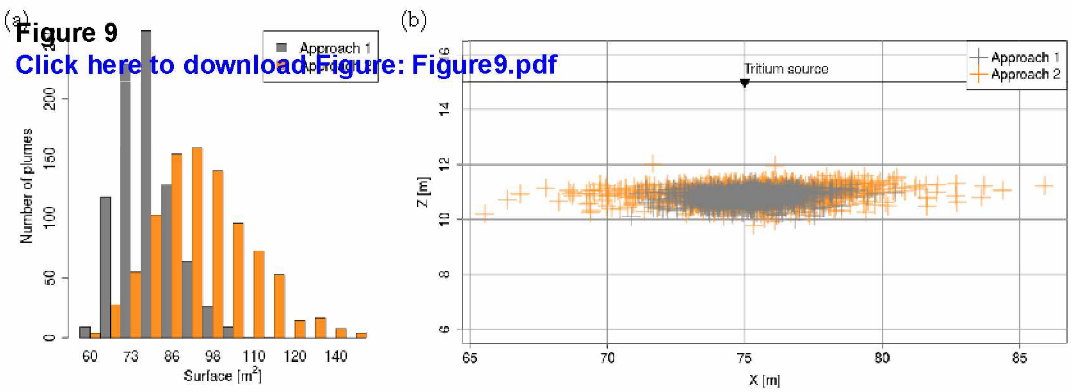

Figure 10
Click here to download Figure: Figure10.pdf