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

Léa Pannecoucke[1], Mathieu Le Coz[2], Xavier Freulon[1], Chantal de Fouquet[1]

[1] MINES ParisTech, PSL University, Centre de Géosciences
[2] Institut de Radioprotection et de Sûreté Nucléaire, PSE-ENV/SEDRE

AquaConSoil Antwerp 2019

Aquaconsoil 2019 – 20-24 May – Antwerp
New digital and management developments

Context

How to **characterize contamination in soils or groundwater** when dealing with a polluted site or an industrial site needing remediation and with a small amount of available observations?

Context

How to **characterize contamination in soils or groundwater** when dealing with a polluted site or an industrial site needing remediation and with a small amount of available observations?

Classical approaches:

1. **Geostatistical estimation (kriging)**
   - \+ Observations of activity are honored
   - \- Physical information is not taken into account
   - \- Performances are limited if only a few observations are available

Context

How to **characterize contamination in soils or groundwater** when dealing with a polluted site or an industrial site needing remediation and with a small amount of available observations?

Classical approaches:

1. **Geostatistical estimation (kriging)**
   - \+ Observations of activity are honored
   - \- Physical information is not taken into account
   - \- Performances are limited if only a few observations are available
2. **Flow and transport simulations**
   - \+ Physically-based simulations
   - \- Uncertainties in modeling parameters
   - \- Observations of activity are not honored (unless inverse modeling is performed)

Context

How to **characterize contamination in soils or groundwater** when dealing with a polluted site or an industrial site needing remediation and with a small amount of available observations?

Classical approaches:

1. **Geostatistical estimation (kriging)**
   - \+ Observations of activity are honored
   - \- Physical information is not taken into account
   - \- Performances are limited if only a few observations are available
2. **Flow and transport simulations**
   - \+ Physically-based simulations
   - \- Uncertainties in modeling parameters
   - \- Observations of activity are not honored (unless inverse modeling is performed)

Development of a method using physical information given by physically-based simulation outputs into a geostatistical framework.

Presentation outline

1. Kriging with numerical variograms

2. A synthetic reference test case

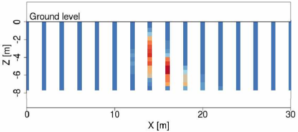
3. Comparison of kriging with numerical variograms to classical krigings

Kriging with numerical variograms

Kriging with numerical variograms

Limits of the experimental variogram computed from observations

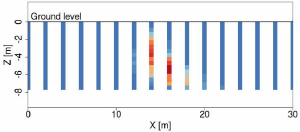
Numerous observations (boreholes)



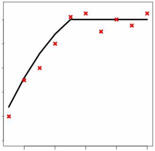
Kriging with numerical variograms

Limits of the experimental variogram computed from observations

Numerous observations (boreholes)



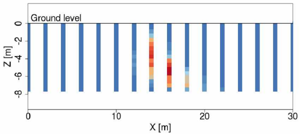
Variogram fitting



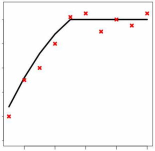
Kriging with numerical variograms

Limits of the experimental variogram computed from observations

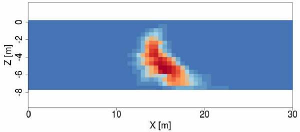
Numerous observations (boreholes)



Variogram fitting



Kriging

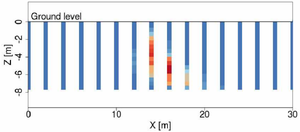


Classical approach

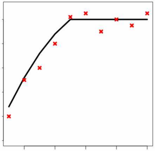
Kriging with numerical variograms

Limits of the experimental variogram computed from observations

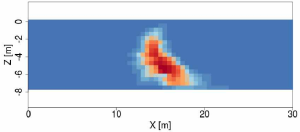
Numerous observations (boreholes)



Variogram fitting

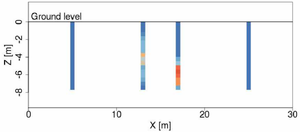


Kriging



Classical approach

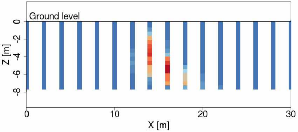
Small amount of observations



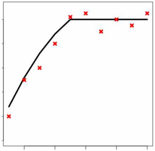
Kriging with numerical variograms

Limits of the experimental variogram computed from observations

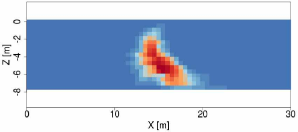
Numerous observations (boreholes)



Variogram fitting

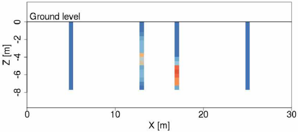


Kriging

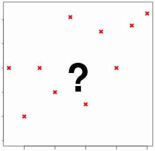


Classical approach

Small amount of observations



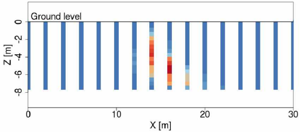
Variogram fitting



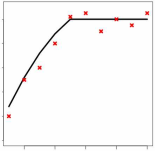
Kriging with numerical variograms

Limits of the experimental variogram computed from observations

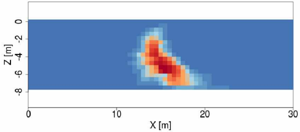
Numerous observations (boreholes)



Variogram fitting

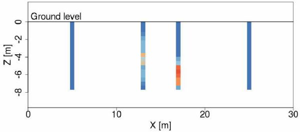


Kriging

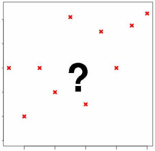


Classical approach

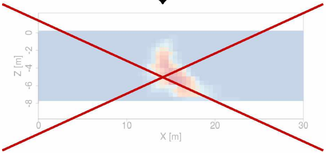
Small amount of observations



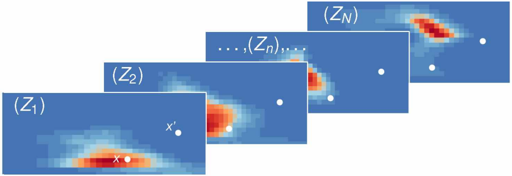
Variogram fitting



Kriging ?



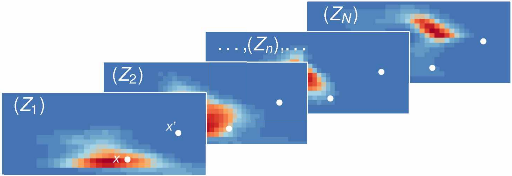
~~Classical approach~~



N: number of simulations
(x,x'): couple of points ranging the whole domain

Kriging with numerical variograms

Principle of the method



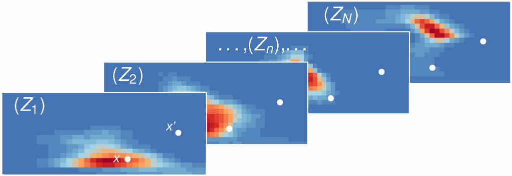
N: number of simulations
(x,x'): couple of points ranging the whole domain

Numerical variograms computed from simulation outputs

$$\gamma(x,x') = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{2}[Z_n(x) - Z_n(x')]^2$$

Kriging with numerical variograms

Principle of the method

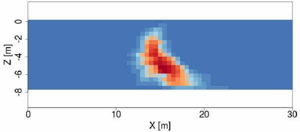


N: number of simulations
(x,x'): couple of points ranging the whole domain

Numerical variograms computed from simulation outputs

$$\gamma(x,x') = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{2}[Z_n(x) - Z_n(x')]^2$$

Kriging



Kriging with numerical variograms (KNV)

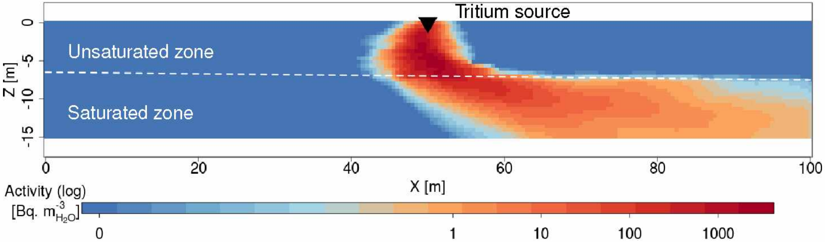
A synthetic reference test case

Synthetic reference test case (2D)

Synthetic aquifer (vertical section):

- single sandy facies with **heterogeneous textural properties** (sand, silt, clay contents);
- fixed head and hydraulic gradient imposed;
- unsaturated zone $\sim$ 7 m deep;
- time variable percolation flow;
- contamination due to a point source of tritium.

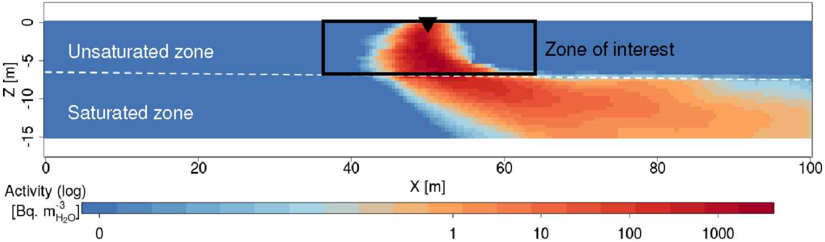
Reference tritium plume (4 years after tritium injection) simulated with the flow and transport simulation code MELODIE:



Synthetic aquifer (vertical section):

- single sandy facies with **heterogeneous textural properties** (sand, silt, clay contents);
- fixed head and hydraulic gradient imposed;
- unsaturated zone $\sim$ 7 m deep;
- time variable percolation flow;
- contamination due to a point source of tritium.

Reference tritium plume (4 years after tritium injection) simulated with the flow and transport simulation code MELODIE:

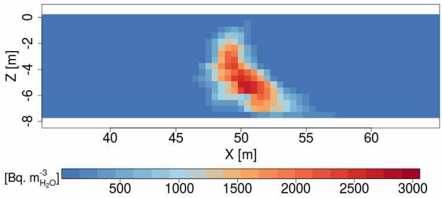


Synthetic reference test case (2D)

Synthetic aquifer (vertical section):

- single sandy facies with **heterogeneous textural properties** (sand, silt, clay contents);
- fixed head and hydraulic gradient imposed;
- unsaturated zone ∼ 7 m deep;
- time variable percolation flow;
- contamination due to a point source of tritium.

Reference tritium plume (4 years after tritium injection) simulated with the flow and transport simulation code MELODIE:



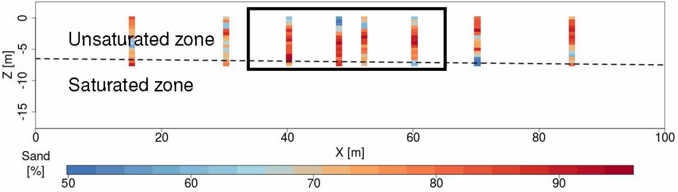
Zoom-in the zone of interest (unsaturated zone)

and non-log color scale.

Available observations

Sampling of the reference test case along vertical boreholes

Observations of textural properties (sand, silt, clay contents).

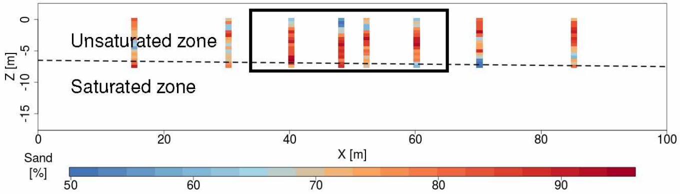


Example with sand proportion.

Available observations

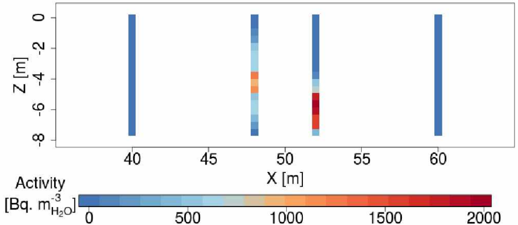
Sampling of the reference test case along vertical boreholes

Observations of textural properties (sand, silt, clay contents).



Example with sand proportion.

Observations of tritium activities.

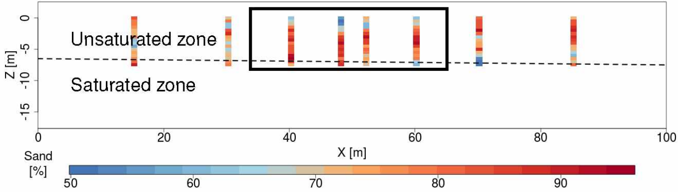


From the observations of activities:
how to estimate tritium activity on the whole domain using kriging with numerical variograms?

Available observations

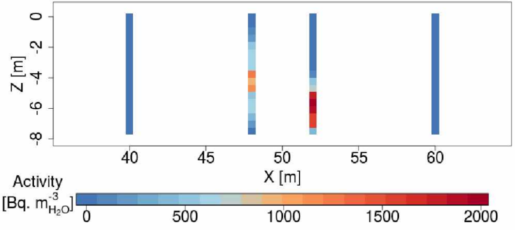
Sampling of the reference test case along vertical boreholes

Observations of textural properties (sand, silt, clay contents).



Example with sand proportion.

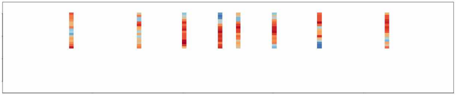
Observations of tritium activities.



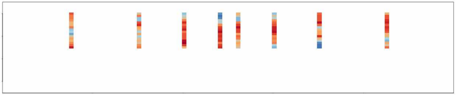
From the observations of activities:
how to estimate tritium activity on the whole domain using kriging with numerical variograms?

$\Longrightarrow$ Building a set of simulations representing the spatial variability of the contaminant plume, using observations of textural properties.

Observations of textures (sand, silt, clay contents)



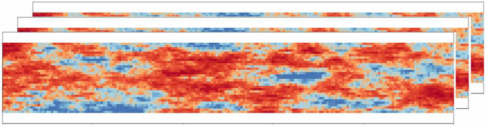
Observations of textures (sand, silt, clay contents)



Geostatistical simulations

(R/RGeostats)

Conditionnal simulations of textures

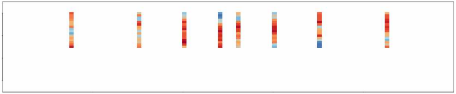


($\times$1000)

Building of a set of simulations

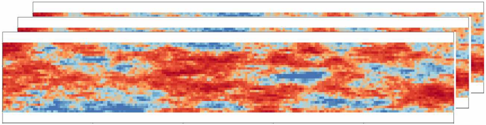
Textural approach

Observations of textures (sand, silt, clay contents)



Geostatistical simulations (R/RGeostats)

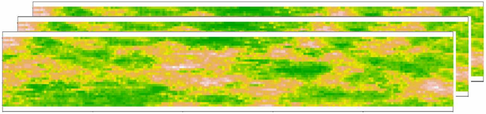
Conditionnal simulations of textures



(×1000)

Pedotransfer function

Conditionnal simulations of hydraulic parameters (Mualem-van Genuchten)

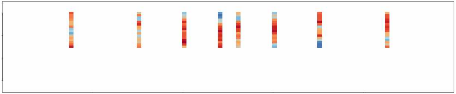


(×1000)

Building of a set of simulations

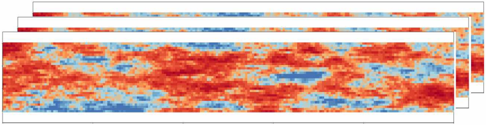
Textural approach

Observations of textures (sand, silt, clay contents)



Geostatistical simulations (R/RGeostats)

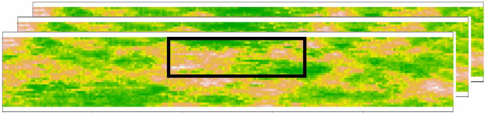
Conditionnal simulations of textures



($\times$1000)

Pedotransfer function

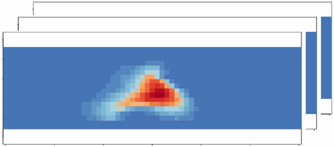
Conditionnal simulations of hydraulic parameters
(Mualem-van Genuchten)



($\times$1000)

MELODIE code

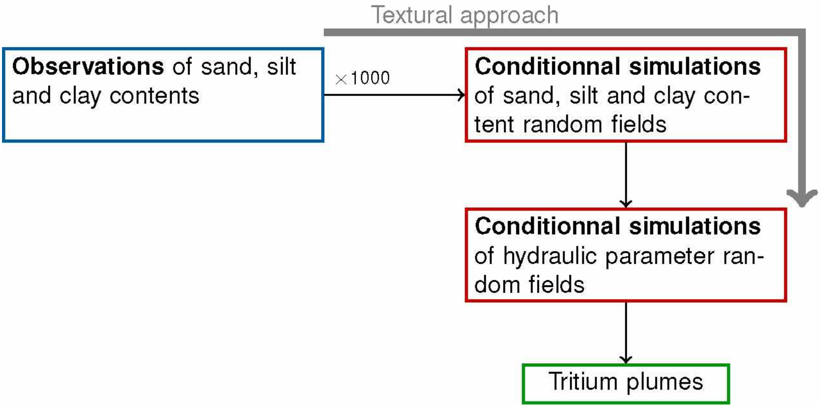
Simulations of tritium plumes
(at a given time step)



($\times$1000)

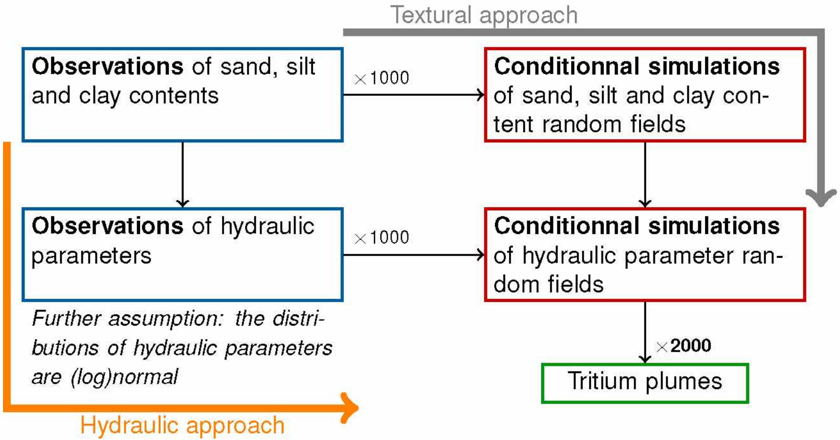
Building of a set of simulations

Textural approach *vs.* hydraulic approach



Building of a set of simulations

Textural approach *vs.* hydraulic approach



Building of a set of simulations

Textural approach *vs.* hydraulic approach

Textural approach

Observations of sand, silt and clay contents $\xrightarrow{\times 1000}$ **Conditionnal simulations** of sand, silt and clay content random fields

Observations of hydraulic parameters $\xrightarrow{\times 1000}$ **Conditionnal simulations** of hydraulic parameter random fields

Further assumption: the distributions of hydraulic parameters are (log)normal

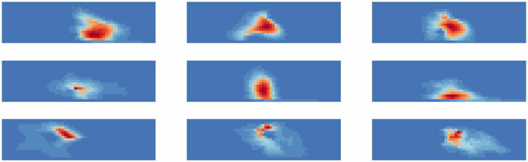
$\times$**2000** → Tritium plumes

Hydraulic approach

The other input parameters (boundary conditions, initial conditions) are fixed.

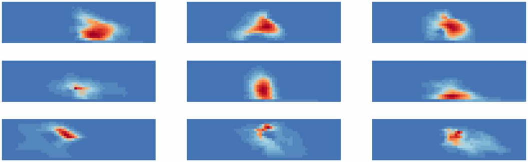
Focus on the uncertainties due to the **spatial variability of hydraulic parameters.**

Simulated plumes



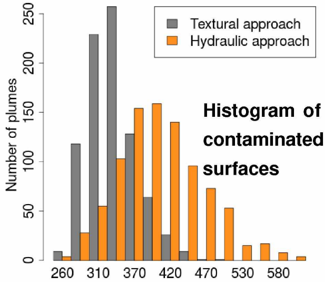
9 examples of simulated tritium plumes in the unsaturated zone

Simulated plumes

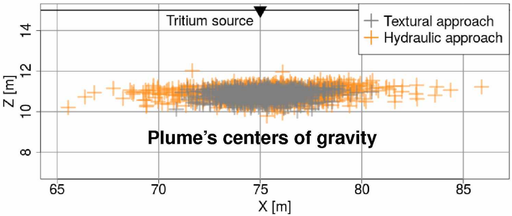


9 examples of simulated tritium plumes in the unsaturated zone

Diversity among the simulated plumes (surfaces and locations):



Histogram of contaminated surfaces



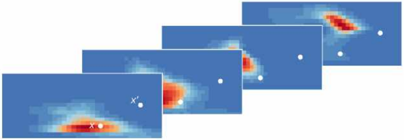
Plume's centers of gravity

The characterization of the tritium plume within the unsaturated zone is not accurate, due to the **uncertainties related to hydraulic parameters**.

Reminder: kriging with numerical variograms

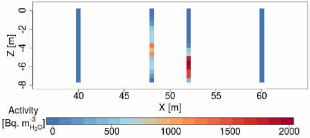
Objective: estimating the tritium activity on the whole domain using observations from a few boreholes.

Method

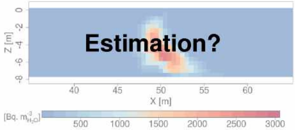


Numerical variograms $\gamma(x, x')$

+



Observations of activity from the reference

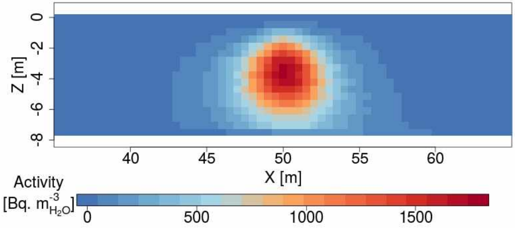


Comparison of kriging with numerical variograms to classical krigings

Two benchmark methods:

- **Ordinary kriging (OK)**, which is widely used but known to perform poorly when the number of data is too small or when the phenomenon under study is complex;
- **Kriging with an external drift (KED)**, which enables the incorporation of auxiliary variables to take non-stationarity into account.

  The auxiliary variable is the empirical mean of the set of simulations.

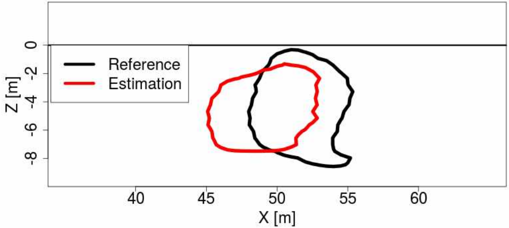


Performance indicators

- **Maps** of estimation, kriging error and kriging error standard deviation;
- Mean absolute error (**MAE**), root mean square error (**RMSE**), mean relative error (**MRE**) over the zone of interest (unsaturated zone);

Performance indicators

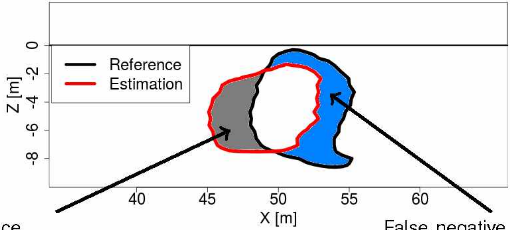
- **Maps** of estimation, kriging error and kriging error standard deviation;
- Mean absolute error (**MAE**), root mean square error (**RMSE**), mean relative error (**MRE**) over the zone of interest (unsaturated zone);
- For given pollution thresholds z, the **false-positive and false-negative** surfaces divided by the contaminated surface on the reference.



Delimitation of the plumes for a given threshold z.

Performance indicators

- **Maps** of estimation, kriging error and kriging error standard deviation;
- Mean absolute error (**MAE**), root mean square error (**RMSE**), mean relative error (**MRE**) over the zone of interest (unsaturated zone);
- For given pollution thresholds z, the **false-positive and false-negative** surfaces divided by the contaminated surface on the reference.

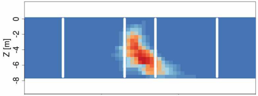


False positive surface
(extra remediation costs)

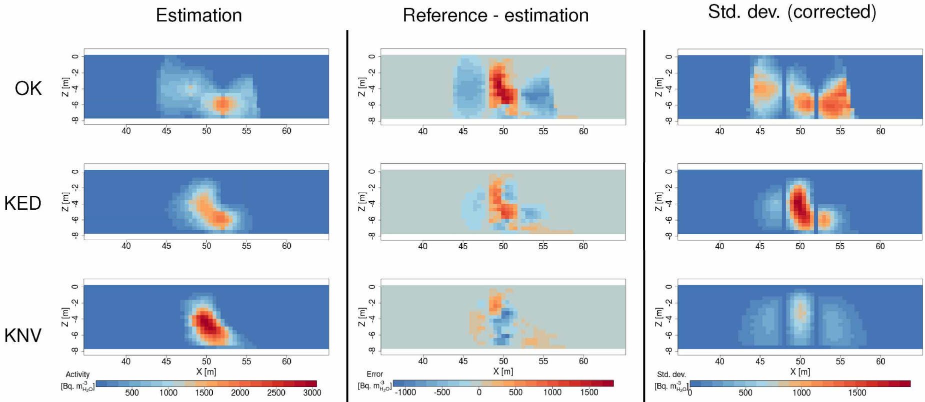
False negative surface
(health and environmental hazard)

Maps

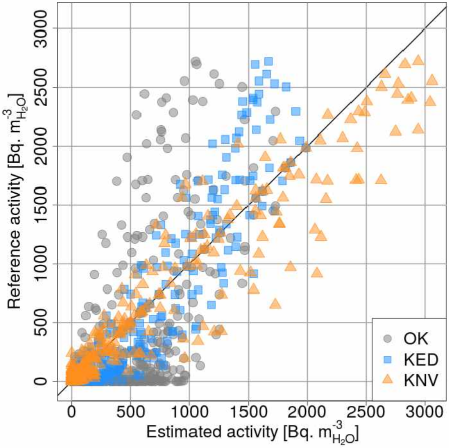
Results (1/3)



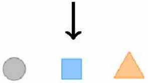
Reference tritium plume



Correlation plot reference *vs.* estimation

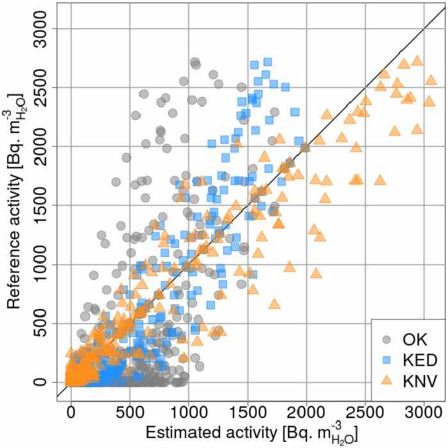


1 point of the domain

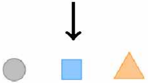


OK and KED underestimate high values of activity while KNV overestimates those values.

Correlation plot reference *vs.* estimation



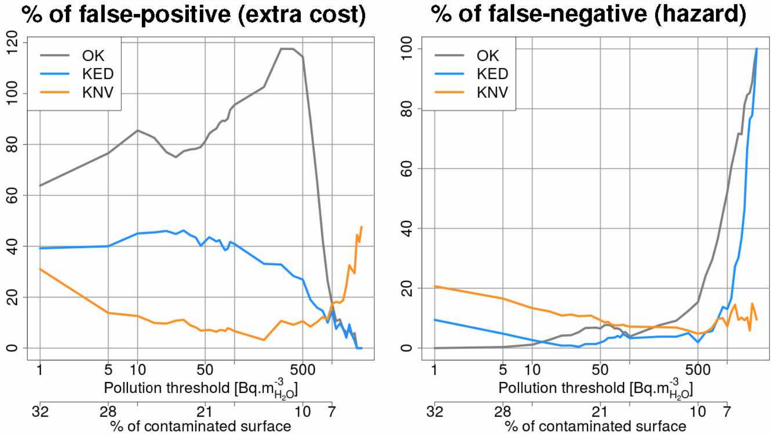
1 point of the domain



Global statistics

	OK	KED	KNV
MAE [Bq.m^{-3}$_{H2O}$]	173	71	**47**
RMSE[Bq.m^{-3}$_{H2O}$]	348	174	**147**
MRE [-]	-47	-6.8	**-0.8**

OK and KED underestimate high values of activity while KNV overestimates those values. Global errors are smaller for KNV than for KED and OK.

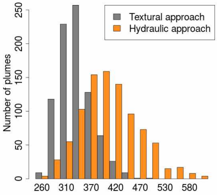


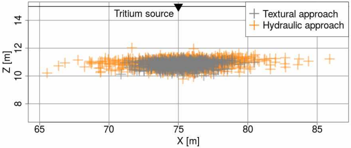
KNV leads to smaller false-positive surfaces. For $z > 500 Bq.m^{-3}_{H2O}$, KNV leads to smaller false-negative surfaces.

The standard deviation of kriging error is not taken into account for the classification.

Conclusions and outlooks

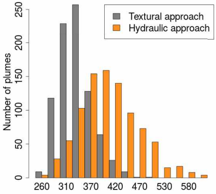
- The characterization of the tritium plume within the unsaturated zone is not accurate, because of the **uncertainties related to hydraulic parameters**, and even if the initial and boundary conditions of the flow and transport model are fixed.

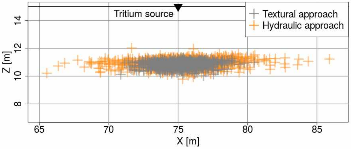




Conclusions and outlooks

- The characterization of the tritium plume within the unsaturated zone is not accurate, because of the **uncertainties related to hydraulic parameters**, and even if the initial and boundary conditions of the flow and transport model are fixed.



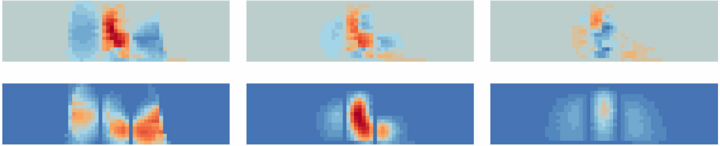


Other input parameters also impact flow and solute transport in the unsaturated zone. Is it interesting to take those uncertainties into account?

Conclusions and outlooks

- The characterization of the tritium plume within the unsaturated zone is not accurate, due to the **uncertainties related to hydraulic parameters**, even if the initial and boundary conditions of the flow and transport model are fixed.

- **KNV improves the estimates of tritium plumes in the unsaturated zone** compared to OK or KED: estimation errors and standard deviation of errors are reduced.



Conclusions and outlooks

- The characterization of the tritium plume within the unsaturated zone is not accurate, due to the **uncertainties related to hydraulic parameters**, even if the initial and boundary conditions of the flow and transport model are fixed.
- **KNV improves the estimates of tritium plumes in the unsaturated zone** compared to OK or KED: estimation errors and standard deviation of errors are reduced.
- KNV is even more interesting **when the number of observations of pollutant concentration is reduced**. It also works when the boreholes are located around the zone of high values of activities and not in its middle.

  *With 7 boreholes: MAE reduced of **52%** compared to OK.*
  *With 4 boreholes: MAE reduced of **73%** compared to OK.*

Conclusions and outlooks

- The characterization of the tritium plume within the unsaturated zone is not accurate, due to the **uncertainties related to hydraulic parameters**, even if the initial and boundary conditions of the flow and transport model are fixed.

- **KNV improves the estimates of tritium plumes in the unsaturated zone** compared to OK or KED: estimation errors and standard deviation of errors are reduced.

- KNV is even more interesting **when the number of observations of pollutant concentration is reduced**. It also works when the boreholes are located around the zone of high values of activities and not in its middle.

- **Next step: implementation on a real study case.**
- **Generalization to a spatio-temporal framework.**

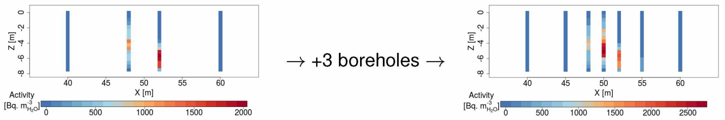
Thank you for your attention.

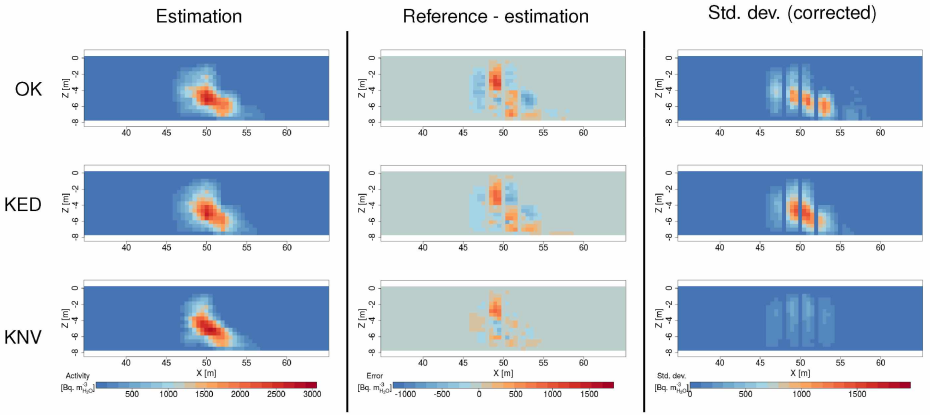
This study is part of *Kri-Terres* project, supported by the French National Radioactive Waste Management Agency (ANDRA) under the French "Investments for the Future" Program.

Back-up slides

Results (7 instead of 4 boreholes)

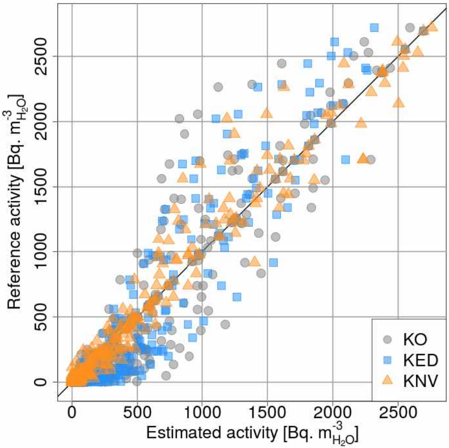
Maps





Results (7 instead of 4 boreholes)

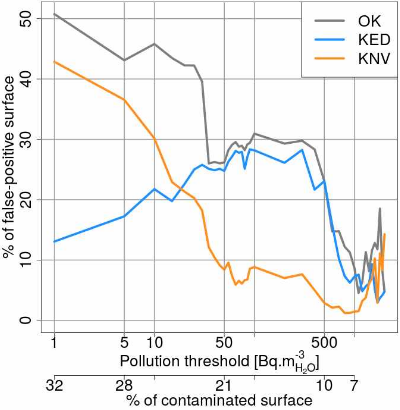
Errors

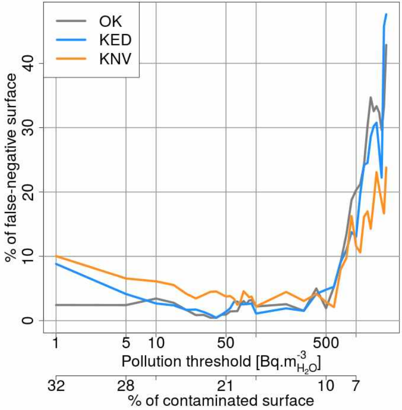


	OK	KED	KNV
MAE	61	53	**29**
RMSE	161	138	**89**
MRE	-4.6	-2.8	**-2.2**

Results (7 instead of 4 boreholes)

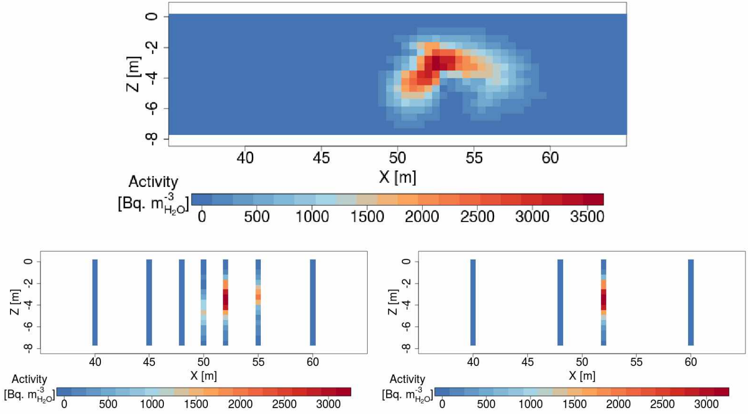
False-positive and false-negative surfaces





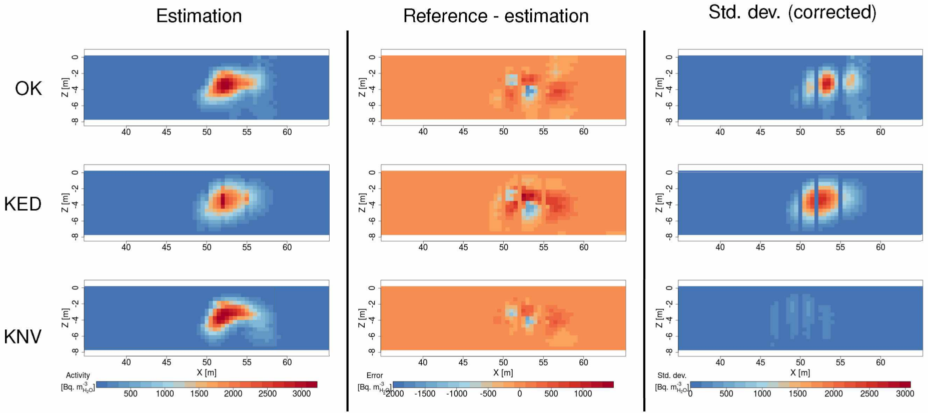
Additional test case

Additional test case, built from the same model but different sand, silt and clay content random fields:



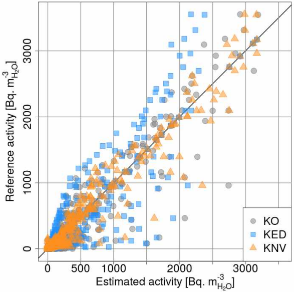
Results for the additional test case (7 boreholes)

Maps

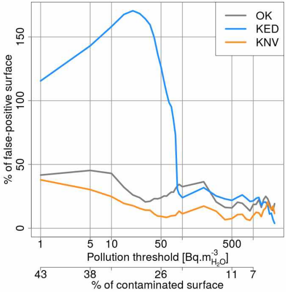


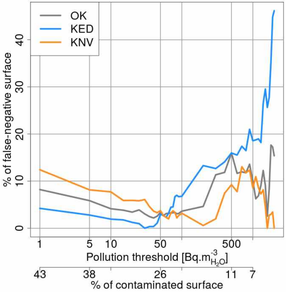
Results for the additional test case (7 boreholes)

Errors and surfaces



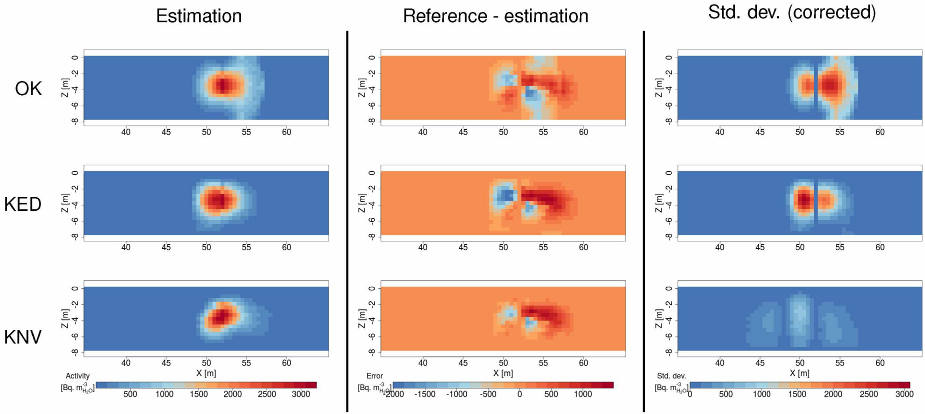
	OK	KED	KNV
MAE	72	119	43
RMSE	184	230	125
MRE	-4.8	-31	-2.2





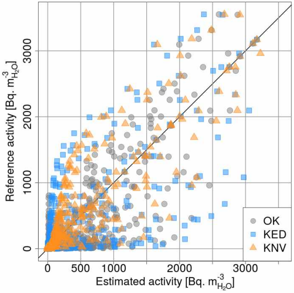
Results for the additional test case (4 boreholes)

Maps

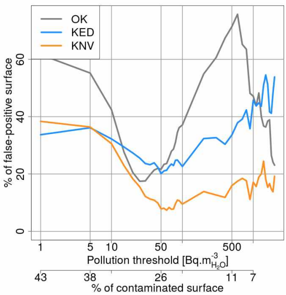


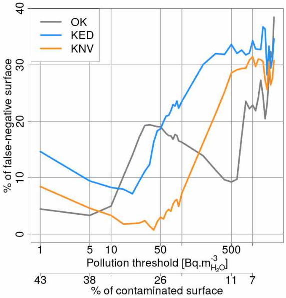
Results for the additional test case (4 boreholes)

Errors and surfaces



	OK	KED	KNV
MAE	139	140	82
RMSE	302	355	233
MRE	-5.6	-4.2	-2.7





Textural approach *vs.* hydraulic approach

to build the hydraulic parameter fields

	Reference test case		Additional test case	
	7 boreholes	4 boreholes	7 boreholes	4 boreholes
KNV-1	30	58	163	484
KNV-2	32	41	44	92
KNV	29	47	43	82

KNV performs better when the hydraulic parameter fields are generated from interpolation of punctual estimate of these parameters (KNV-2) than from conversion of soil texture fields (KNV-1).

The hydraulic approach leads to a higher variability in hydraulic parameter fields and thus in more variable simulated plumes as outputs of the flow and transport model.

Uncertainties in the input parameters

Variogram parameters randomization

The parameters (sill C, horizontal range s_x and vertical range s_z) of the variogram models used to simulate random fields of hydraulic parameters are randomized.

- $C \sim \mathcal{N}(C^{id}, \frac{0.2}{3}C^{id})$ (such that almost all realizations belong to the interval $[0.8C^{id};1.2C^{id}]$).
- $s_z \sim \mathcal{N}(s_z^{id}, \frac{0.2}{3}s_z^{id})$;
- $s_x \sim \mathcal{T}(min = 2s_z^{id}, max = 10s_z^{id}, mode = s_x^{id})$;
- 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).