Université Gustave Eiffel

UNIVERSITÉ Clermont Auvergne

INRAE la science pour la vie, l'humain, la terre
UR REVERSAAL

Data assimilation: overview and constraints for agro-ecology modelling

Claire Lauvernet[1], Claudio Paniconi[2], Laura Gatel[2], Emilie Rouzies[1]
1 INRAE RIVERLY, Lyon-Villeurbanne
2 INRS & Univ. Laval, Québec, Canada.

Séminaire National Fusion d'informations :
Applications aux géosciences et au génie-civil
Le 01 et 02 février 2021

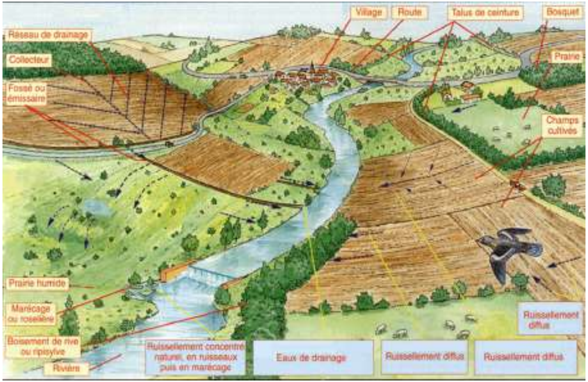
Avec le soutien de :
AUGC
ASSOCIATION UNIVERSITAIRE DE GÉNIE CIVIL

Plan

Context : How to improve the water quality ?

⇒ a better understanding of water and pesticide transfer in soil

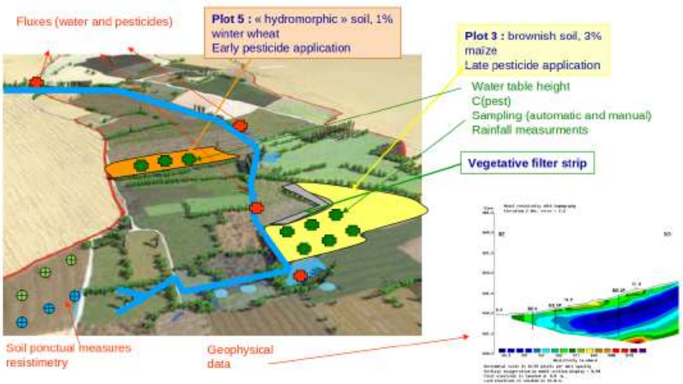
- Spatial heterogeneity of the soils, at all scales
- Soil and agricultural practices are more and more diverse
- Processes that drive the pesticide fate at the catchment scale are complex :
  - Hydrological transfer
  - adsorption
  - degradation



Spatially and temporally heterogeneous data. . .

Availability, quality, quantity of data are heterogeneous in space and time :

- remote sensing images
- field data (lysimeters in soil, water table and river measurments)
- geophysical data

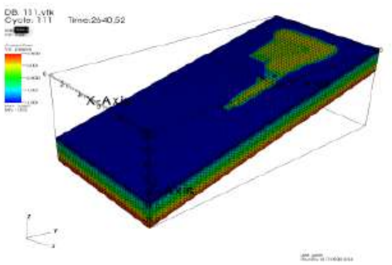


BUT without heavy experiments, this is very difficult to get the pesticides dynamics

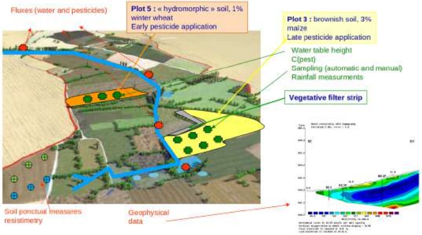
Spatially and temporally heterogeneous data...
...and pesticides modeling at several scales and several complexity degrees

- based on non linear equations and/or conceptual
- unknown boundary and initial conditions
- a large set of spatialized parameters that are difficult to measure/estimate
- many processes affecting pesticide transfer are not (well) represented (e.g., pref. flows)

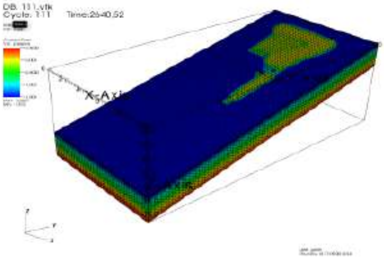
⇒ a high uncertainty (when we it is considered !)



Spatially heterogeneous data. . .



. . . and spatialized modeling

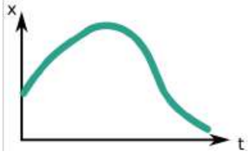


⇒ merging information from the available data and from the model
to get as close as possible to the "true" state

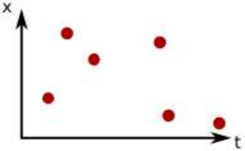
Data Assimilation techniques (or *model-data fusion*)

Plan

- the systematic use of data to constrain a numerical model
- first used in the 1960s in numerical weather forecasting models for short-term predictions of meteorological conditions
- in the 1970s, development in numerical ocean general circulation models (OGCMs)
- poorly developed in other domains (hydrology)

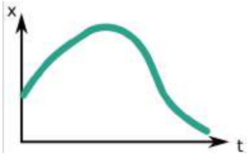


Model

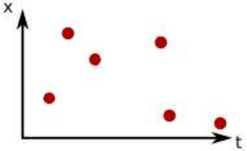


Observations

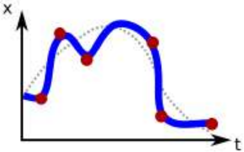
- the systematic use of data to constrain a numerical model
- first used in the 1960s in numerical weather forecasting models for short-term predictions of meteorological conditions
- in the 1970s, development in numerical ocean general circulation models (OGCMs)
- poorly developed in other domains (hydrology)



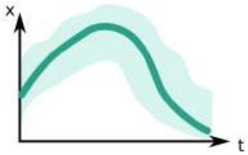
Model



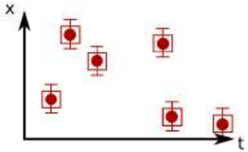
Observations



- the systematic use of data to constrain a numerical model
- first used in the 1960s in numerical weather forecasting models for short-term predictions of meteorological conditions
- in the 1970s, development in numerical ocean general circulation models (OGCMs)
- poorly developed in other domains (hydrology)

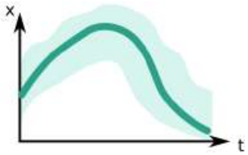


Model

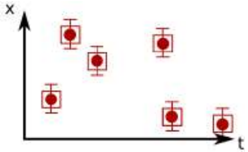


Observations

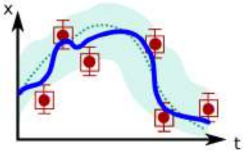
- the systematic use of data to constrain a numerical model
- first used in the 1960s in numerical weather forecasting models for short-term predictions of meteorological conditions
- in the 1970s, development in numerical ocean general circulation models (OGCMs)
- poorly developed in other domains (hydrology)



Model



Observations



Data assimilation

The ingredients

$\mathbf{x} = (\mathbf{x_0},\mathbf{x_1},\ldots,\mathbf{x_N})^\mathbf{T}$ represents the state of system:
streamflow at the outlet, soil moisture, dissolved oxygen concentration in the river, etc.
We don't know it, but we do have information from :

- **the dynamical model** $x_k = M_{k-1\rightarrow k}[x_{k-1}, param] + \eta_k$

  η_k the model error of covariance matrix P_k

INRAØ

Data assimilation

The ingredients

$\mathbf{x} = (\mathbf{x_0}, \mathbf{x_1}, \ldots, \mathbf{x_N})^T$ represents the state of system:
streamflow at the outlet, soil moisture, dissolved oxygen concentration in the river, etc.
We don't know it, but we do have information from :

- **the dynamical model** $x_k = M_{k-1 \to k}[x_{k-1}, param] + \eta_k$

- **the observation model** $y_k = H_k[x_k] + \varepsilon_k$

  y_k is the observation/data at time k
  ε_k the observation error, of covariance matrix R_k, e.g. instrumental error, representativeness
  $H : \mathcal{R}^m \to \mathcal{R}^d$ the observation operator that projects from model space to observational space

Data assimilation

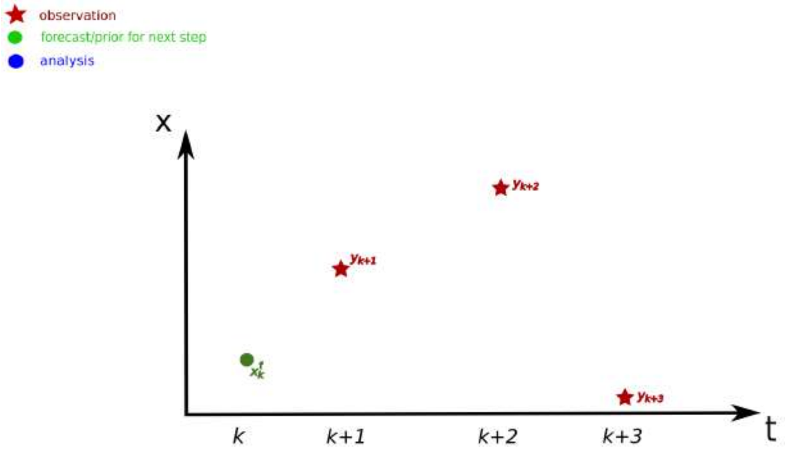
The ingredients

$\mathbf{x} = (\mathbf{x_0}, \mathbf{x_1}, \ldots, \mathbf{x_N})^\mathbf{T}$ represents the state of system:
streamflow at the outlet, soil moisture, dissolved oxygen concentration in the river, etc.
We don't know it, but we do have information from :

- **the dynamical model** $x_k = M_{k-1\to k}[x_{k-1}, param] + \eta_k$
- **the observation model** $y_k = H_k[x_k] + \varepsilon_k$
- We assume that model and obs. errors are random variables $\to$ described by pdf

$\Rightarrow$ Bayesian framework $\Rightarrow$ The Kalman Filter

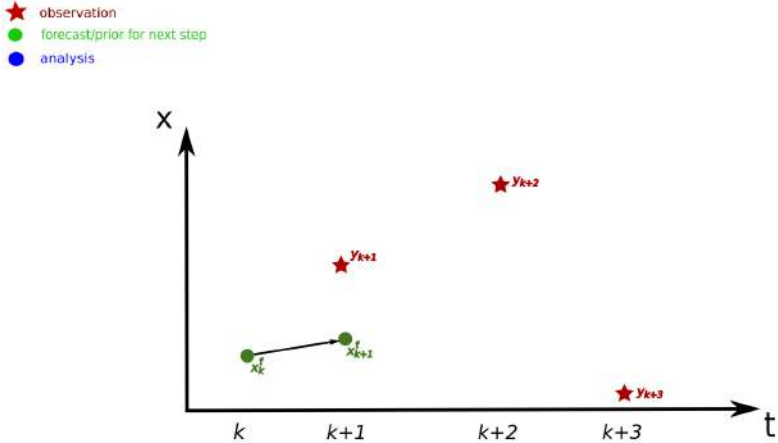
The Kalman Filter (Kalman 1960) : estimate the optimal state at each observation time



0. At time k: $\mathbf{X}_k^f$

The Kalman Filter (Kalman 1960) : estimate the optimal state at each observation

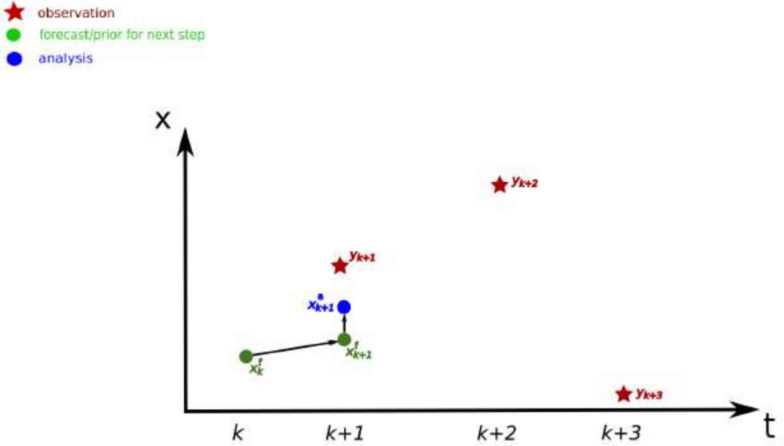
time



1. Forecast step: $\mathbf{X}_{k+1}^f = \mathbf{M}\mathbf{X}_k^f$

The Kalman Filter (Kalman 1960) : estimate the optimal state at each observation

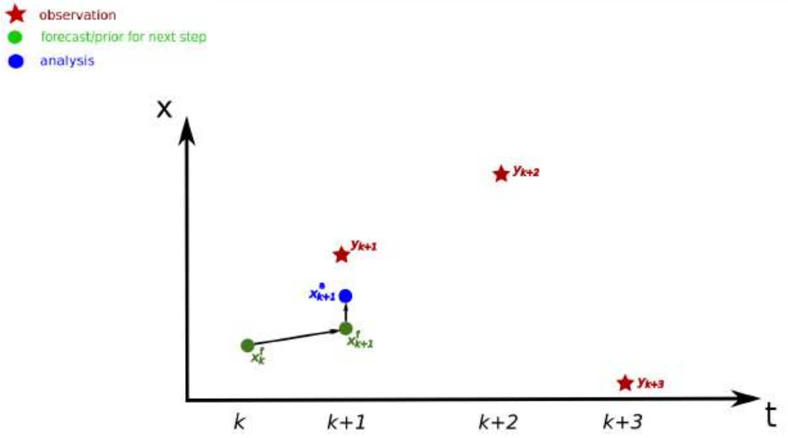
time



2. Analysis step: $\mathbf{X}^a_{k+1} = \mathbf{X}^f_{k+1} + \mathrm{K}_{k+1}(\mathbf{Y}_{k+1} - \mathbf{H}_{k+1}\mathbf{X}^f_{k+1})$

The Kalman Filter (Kalman 1960) : estimate the optimal state at each observation

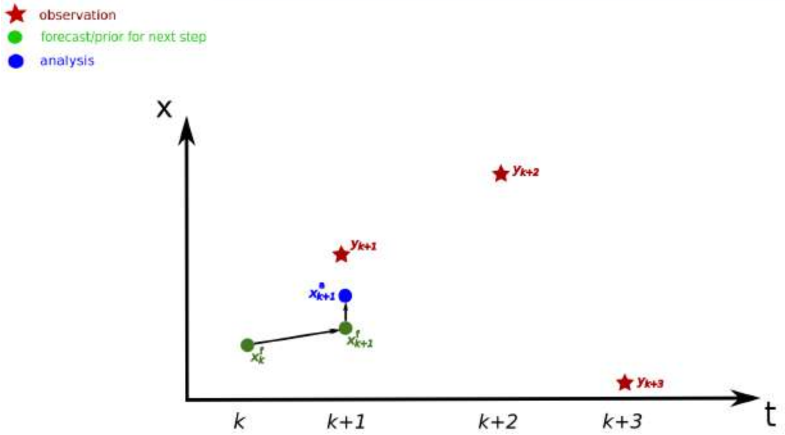
time



2. Analysis step: $\mathbf{X}^a_{k+1} = \mathbf{X}^f_{k+1} + \mathrm{K}_{k+1}(\mathbf{Y}_{k+1} - \mathbf{H}_{k+1}\mathbf{X}^f_{k+1})$

with $\mathrm{K}_{k+1} = \mathbf{P}^f_{k+1}\mathbf{H}^T_{k+1}[\mathbf{H}_{k+1}\mathbf{P}^f_{k+1}\mathbf{H}^T_{k+1} + \mathbf{R}_{k+1}]^{-1}$

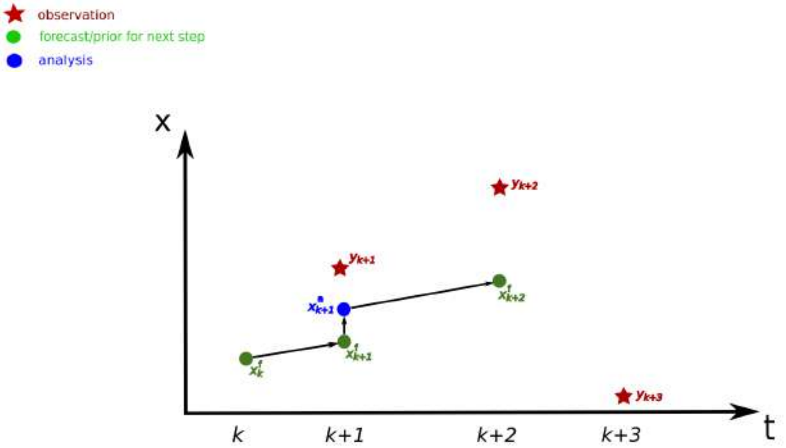
The Kalman Filter (Kalman 1960) : estimate the optimal state at each observation time



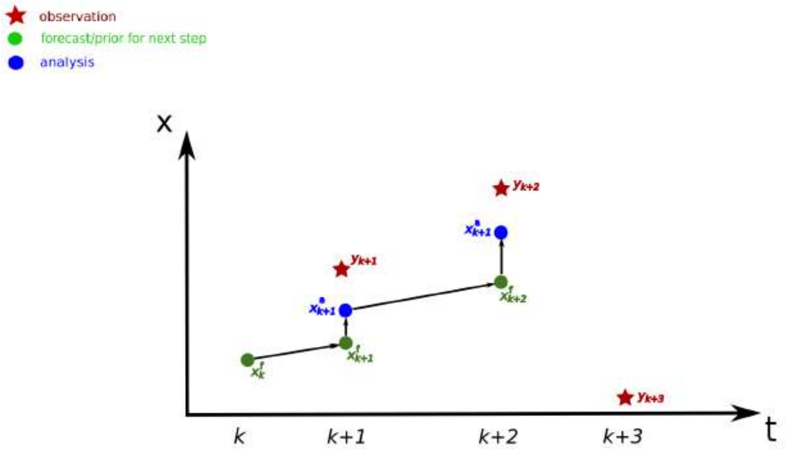
2. Analysis step: $\mathbf{X}^a_{k+1} = \mathbf{X}^f_{k+1} + \mathbf{K}_{k+1}(\mathbf{Y}_{k+1} - \mathbf{H}_{k+1}\mathbf{X}^f_{k+1})$

$$\mathbf{P}^a_{k+1} = \mathbf{P}^f_{k+1} - \mathbf{K}_{k+1}\mathbf{H}_{k+1}\mathbf{P}^f_{k+1}$$

The Kalman Filter (Kalman 1960) : estimate the optimal state at each observation time



The Kalman Filter (Kalman 1960) : estimate the optimal state at each observation time



The Kalman Filter (Kalman 1960) : estimate the optimal state at each observation

time 

The KF assumes that:

- ✓ **all sources of errors are gaussian**
- ✓ **both the observational and dynamical models are linear**

→ not realistic in most cases !

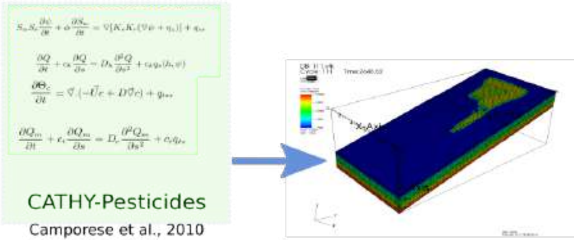
Data assimilation

The method for data assimilation should be suited to spatialized models

- models are physically-based but:
- highly nonlinear equations (Richards, . . .)
- some are more/less conceptual => discontinuities, thresholds

→ **definitely not gaussian !**

→ **Ensemble filter approaches**



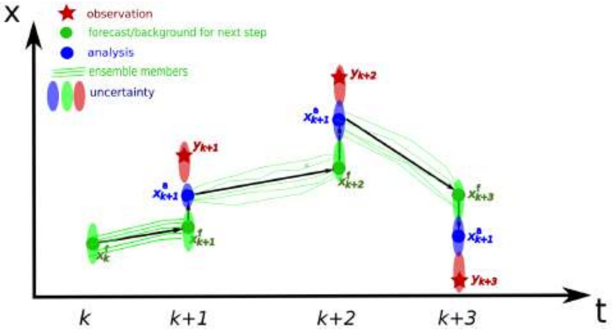
Camporese et al., 2010
Gatel et al., 2019

1

Data assimilation

Ensemble-based methods (Evensen 2003)

- a version of the Kalman filter for nonlinear problems at large dimension
- the state variable distribution is represented by an ensemble of state vectors x_k
- the error covariance matrices are represented by the ensemble covariance

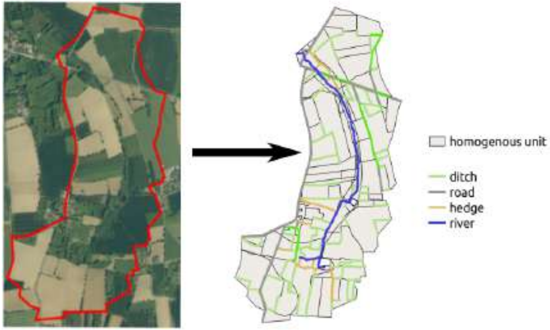


Plan

The PESHMELBA model (Rouzies et al. 2019)

PESticides et Hydrologie: Modélisation à l'EcheLle du BAssin versant

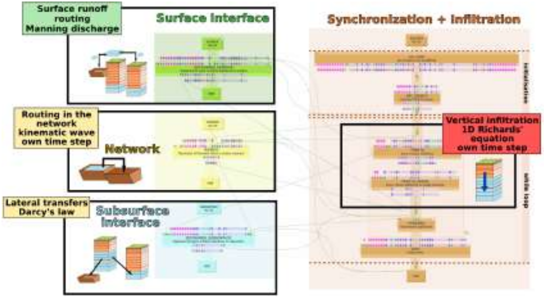
- ✓ Simulation of heterogenous landscapes composed of plots, vegetative filter zones, hedges, ditches and rivers
- ✓ Water transfers on surface and subsurface
- ✓ Solute advection, adsorption and degradation



The PESHMELBA model (Rouzies et al. 2019)

PESticides et Hydrologie: Modélisation à l'EcheLle du BAssin versant

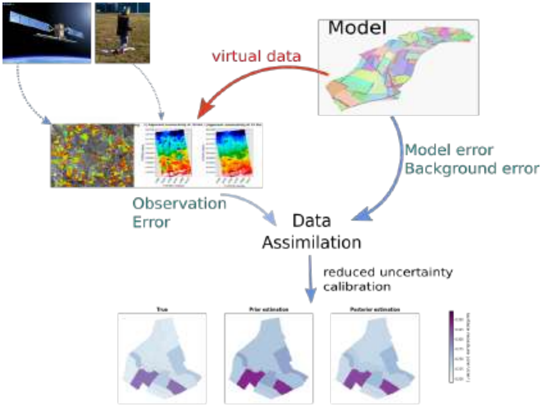
- ✓ Simulation of heterogenous landscapes composed of plots, vegetative filter zones, hedges, ditches and rivers
- ✓ Water transfers on surface and subsurface
- ✓ Solute advection, adsorption and degradation
- ✓ One module ≡ one process or ensemble of processes on a landscape element
- ✓ Coupling of modules within the OpenPALM coupler (Fouilloux and Piacentini 1999) turning the structure flexible



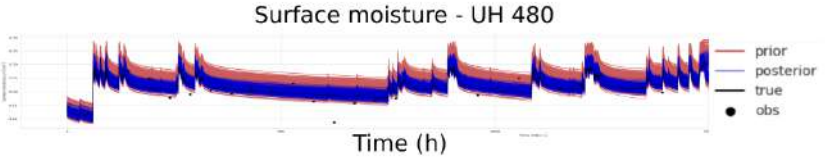
Twin experiments

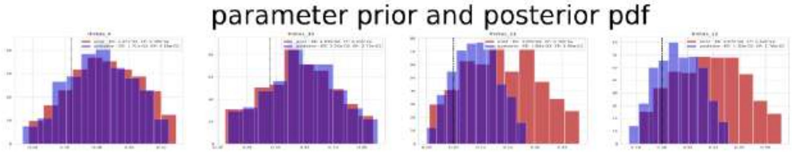
A virtual experiment where we know the true state : an output of the model

- obs $y = x^{true} + error$
- "data" = images of surface soil moisture => virtual data
- model = PESHMELBA
- errors are gaussian



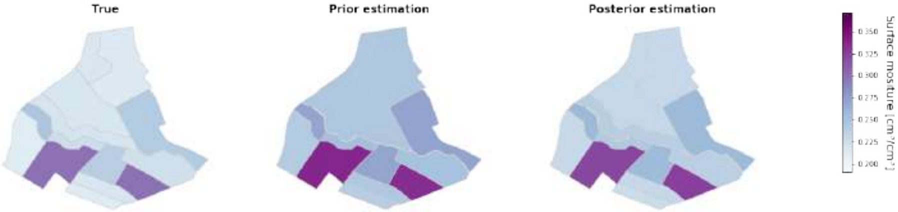
Twin experiments with PESHMELBA





reduction of the uncertainty
some parameters are well estimated

Twin experiments with PESHMELBA



Conclusion

- the twin experiment show the feasibility of DA in pesticide transfer modeling
- assimilate other images than at surface
- define the spatial (and temporal?) correlation in obs. error (E. Rouzies PhD)
- test with real images => scale the model to the data (observation operator)
- combine images with in situ data to improve the water quality predictions
- development on the CATHY model, purely physics-based (less discontinuities?)

⇒ data assimilation = an optimal way to merge information

Thank you!
Any questions ?

Camporese, M. et al. (2010). "Surface-subsurface flow modeling with path-based runoff routing, boundary condition-based coupling, and assimilation of observation data". In: *Water Resources Research* 46.2.

Evensen, Geir (2003). "The Ensemble Kalman Filter: theoretical formulation and practical implementation". In: *Ocean Dynamics* 53.4, pp. 343–367.

Fouilloux, A. and A. Piacentini (1999). "The PALM Project: MPMD Paradigm for an Oceanic Data Assimilation Software". In: *Euro-Par'99 Parallel Processing: 5th International Euro-Par Conference Toulouse, France, August 31 - September 3, 1999 Proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 1423–1430.

Gatel, Laura, Claire Lauvernet, Nadia Carluer, Sylvain Weill, and Claudio Paniconi (2019). "Sobol Global Sensitivity Analysis of a Coupled Surface/Subsurface Water Flow and Reactive Solute Transfer Model on a Real Hillslope". In: *Water* 12.1, p. 121.

Gatel, Laura, Claire Lauvernet, Nadia Carluer, Sylvain Weill, Julien Tournebize, et al. (2019). "Global evaluation and sensitivity analysis of a physically based flow and reactive transport model on a laboratory experiment". In: *Environmental Modelling & Software* 113, pp. 73–83.

Kalman, Rudolph Emil (1960). "A New Approach to Linear Filtering and Prediction Problems". In: *Transactions of the ASME–Journal of Basic Engineering* 82.Series D, pp. 35–45.

Rouzies, Emilie et al. (2019). "From agricultural catchment to management scenarios: A modular tool to assess effects of landscape features on water and pesticide behavior". In: *Science of The Total Environment* 671, pp. 1144–1160.