



Université
Gustave Eiffel



UNIVERSITÉ
Clermont
Auvergne



Data assimilation: overview and constraints for agro-ecology modelling

Claire Lauvernet¹, Claudio Paniconi², Laura Gatel², Emilie Rouzies¹

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

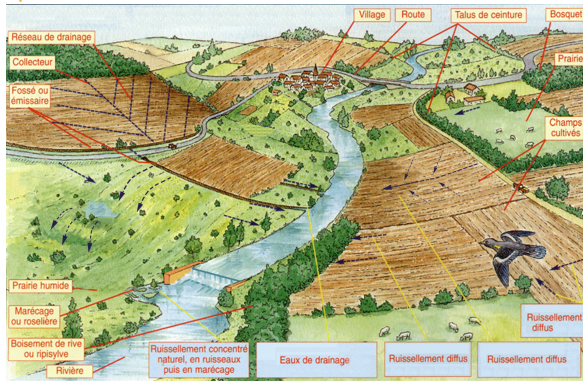
ASSOCIATION UNIVERSITAIRE DE GÉNIE CIVIL

Plan

Context : How to improve the water quality ?

⇒ a better understanding of water and pesticide transfer in soil

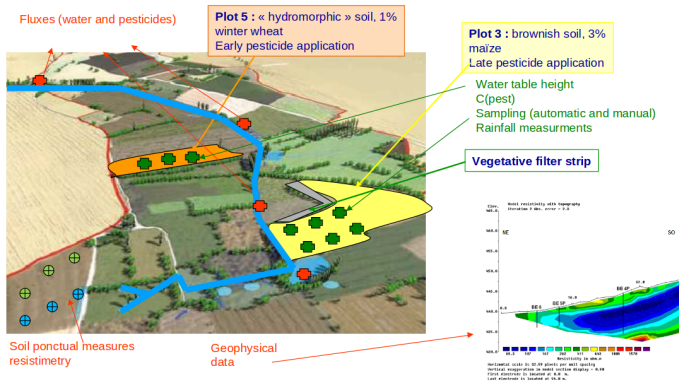
- Spatial heterogeneity of the soils, at all scales
- Soil and agricultural practices are more and more diverse
- Processes that drive the pesticide fate at the catchment scale are complex :
 - Hydrological transfer
 - adsorption
 - degradation



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



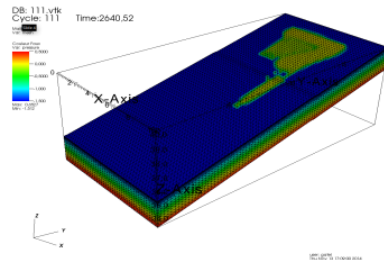
BUT without heavy experiments, this is very difficult to get the pesticides dynamics

Spatially and temporally heterogeneous data...

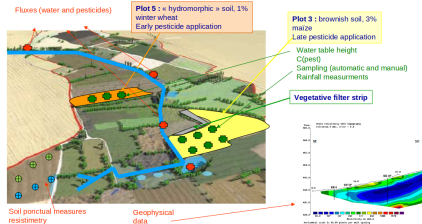
... and pesticides modeling at several scales and several complexity degrees

- 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)

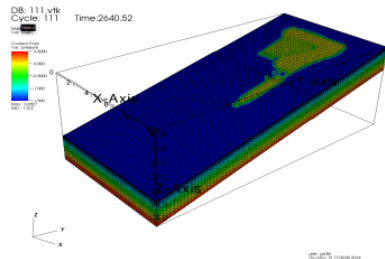
⇒ a high uncertainty (when we it is considered !)



Spatially heterogeneous data...



... and spatialized modeling

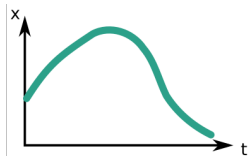
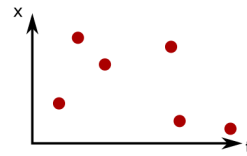


⇒ merging information from the available data and from the model
to get as close as possible to the “true” state

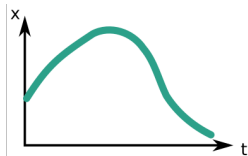
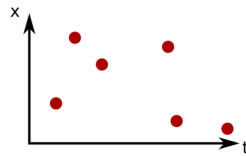
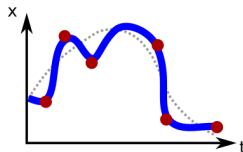
Data Assimilation techniques (or *model-data fusion*)

Plan

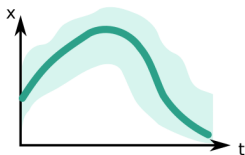
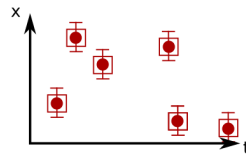
- the systematic use of data to constrain a numerical model
- first used in the 1960s in numerical weather forecasting models for short-term predictions of meteorological conditions
- in the 1970s, development in numerical ocean general circulation models (OGCMs)
- poorly developed in other domains (hydrology)

**Model****Observations**

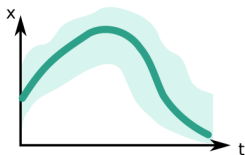
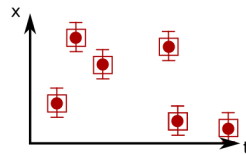
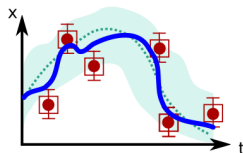
- the systematic use of data to constrain a numerical model
- first used in the 1960s in numerical weather forecasting models for short-term predictions of meteorological conditions
- in the 1970s, development in numerical ocean general circulation models (OGCMs)
- poorly developed in other domains (hydrology)

**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: overview and constraints for agro-ecology modelling

Data assimilation

The ingredients

$\mathbf{x} = (\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_N)^T$ 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** $x_k = M_{k-1 \rightarrow k}[x_{k-1}, param] + \eta_k$

η_k the model error of covariance matrix P_k

Data assimilation

The ingredients

$\mathbf{x} = (x_0, x_1, \dots, x_N)^T$ 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** $x_k = M_{k-1 \rightarrow k}[x_{k-1}, param] + \eta_k$

- **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

$H : \mathcal{R}^m \rightarrow \mathcal{R}^d$ the observation operator that projects from model space to observational space

Data assimilation

The ingredients

$\mathbf{x} = (x_0, x_1, \dots, x_N)^T$ 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 :

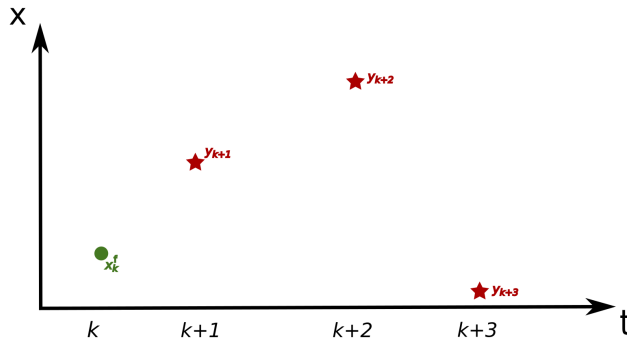
- **the dynamical model** $x_k = M_{k-1 \rightarrow k}[x_{k-1}, param] + \eta_k$
- **the observation model** $y_k = H_k[x_k] + \varepsilon_k$
- We assume that model and obs. errors are random variables \rightarrow described by pdf

\Rightarrow Bayesian framework \Rightarrow The Kalman Filter

The Kalman Filter (Kalman1960) : estimate the optimal state at each observation

- ★ observation
- forecast/prior for next step

time ● analysis

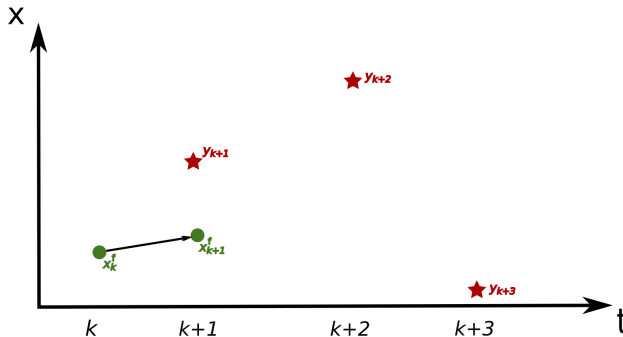


0. At time k : x_k^f

The Kalman Filter (Kalman1960) : estimate the optimal state at each observation

- ★ observation
- forecast/prior for next step

time ● analysis

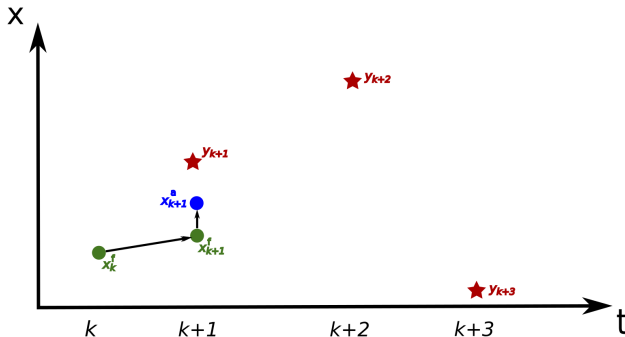


1. Forecast step: $\mathbf{x}_{k+1}^f = \mathbf{M}\mathbf{x}_k^f$

The Kalman Filter (Kalman1960) : estimate the optimal state at each observation

- ★ observation
- forecast/prior for next step

time ● analysis

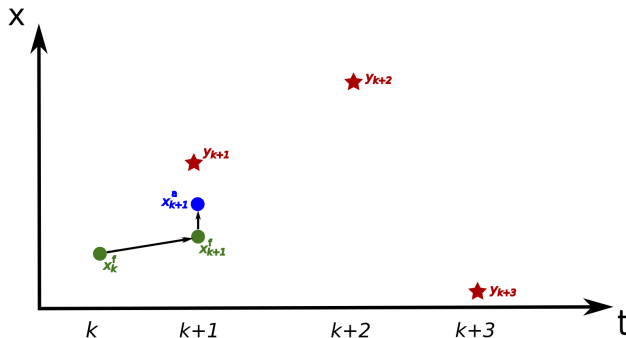


2. Analysis step: $\mathbf{x}_{k+1}^a = \mathbf{x}_{k+1}^f + \mathbf{K}_{k+1}(\mathbf{y}_{k+1} - \mathbf{H}_{k+1}\mathbf{x}_{k+1}^f)$

The Kalman Filter (Kalman1960) : estimate the optimal state at each observation

- ★ observation
- forecast/prior for next step

time ● analysis



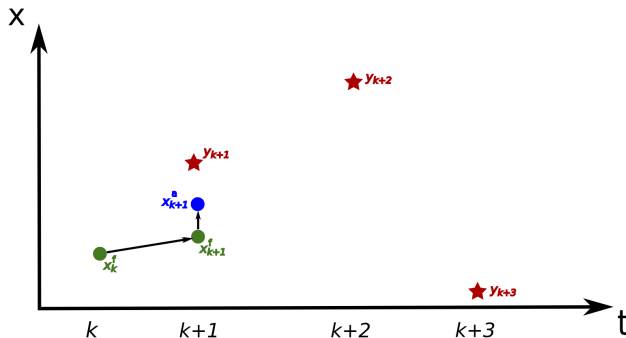
2. Analysis step: $\mathbf{x}_{k+1}^a = \mathbf{x}_{k+1}^f + \mathbf{K}_{k+1}(\mathbf{y}_{k+1} - \mathbf{H}_{k+1}\mathbf{x}_{k+1}^f)$

with $\mathbf{K}_{k+1} = \mathbf{P}_{k+1}^f \mathbf{H}_{k+1}^T [\mathbf{H}_{k+1} \mathbf{P}_{k+1}^f \mathbf{H}_{k+1}^T + \mathbf{R}_{k+1}]^{-1}$

The Kalman Filter (Kalman1960) : estimate the optimal state at each observation

- ★ observation
- forecast/prior for next step

time ● analysis



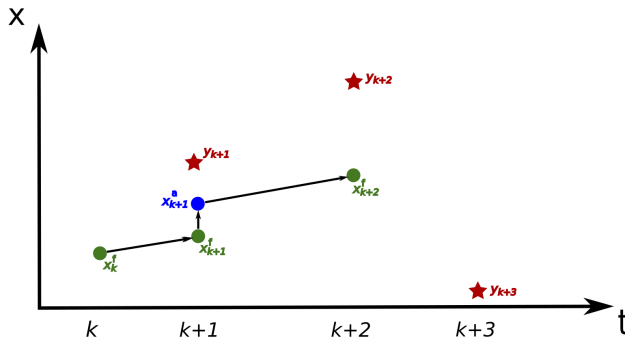
2. Analysis step: $\mathbf{x}_{k+1}^a = \mathbf{x}_{k+1}^f + \mathbf{K}_{k+1}(\mathbf{y}_{k+1} - \mathbf{H}_{k+1}\mathbf{x}_{k+1}^f)$

$$\mathbf{P}_{k+1}^a = \mathbf{P}_{k+1}^f - \mathbf{K}_{k+1}\mathbf{H}_{k+1}\mathbf{P}_{k+1}^f$$

The Kalman Filter (**Kalman1960**) : estimate the optimal state at each observation

- ★ observation
- forecast/prior for next step

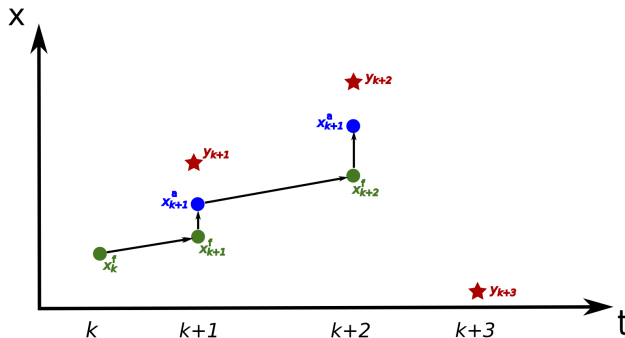
time ● analysis



The Kalman Filter (**Kalman1960**) : estimate the optimal state at each observation

- ★ observation
- forecast/prior for next step

time ● analysis



The Kalman Filter (**Kalman1960**) : estimate the optimal state at each observation

- ★ observation
- forecast/prior for next step
- analysis

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**

1

$$S_w S_s \frac{\partial \psi}{\partial t} + \phi \frac{\partial S_w}{\partial t} = \nabla [K_s K_r (\nabla \psi + \eta_s)] + q_{ss}$$

$$\frac{\partial Q}{\partial t} + c_h \frac{\partial Q}{\partial s} = D_h \frac{\partial^2 Q}{\partial s^2} + c_h q_s(h, \psi)$$

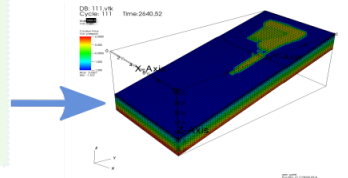
$$\frac{\partial \Theta_c}{\partial t} = \nabla \cdot (-\vec{U}c + D\nabla c) + q_{cs}$$

$$\frac{\partial Q_m}{\partial t} + c_t \frac{\partial Q_m}{\partial s} = D_c \frac{\partial^2 Q_m}{\partial s^2} + c_t q_{ts}$$

CATHY-Pesticides

Camporese et al., 2010

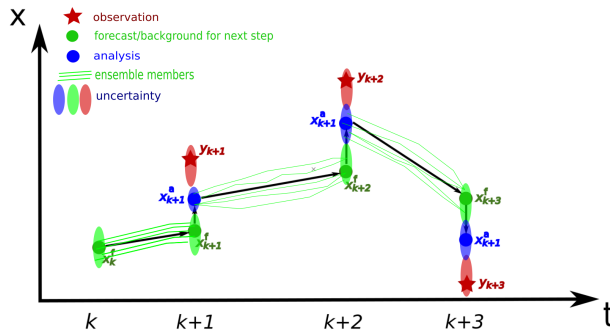
Gatel et al., 2019



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

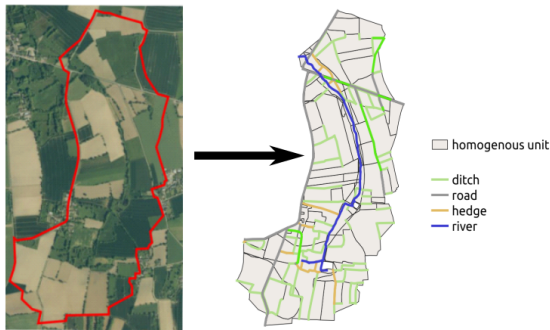


Plan

The PESHMELBA model (Rouzies2019)

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

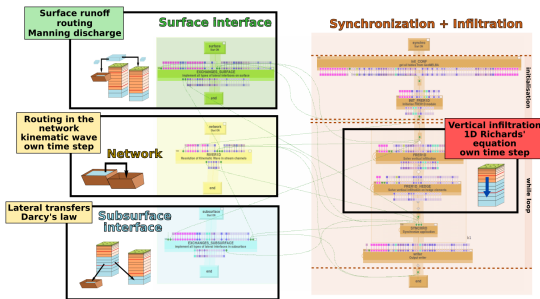
- ✓ Simulation of heterogenous landscapes composed of plots, vegetative filter zones, hedges, ditches and rivers
- ✓ Water transfers on surface and subsurface
- ✓ Solute advection, adsorption and degradation



The PESHMELBA model (**Rouzies2019**)

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

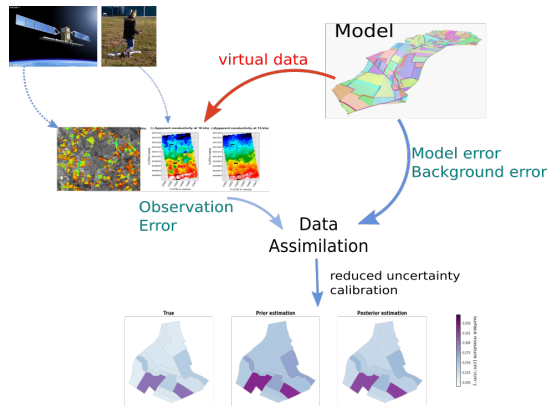
- ✓ Simulation of heterogenous landscapes composed of plots, vegetative filter zones, hedges, ditches and rivers
- ✓ Water transfers on surface and subsurface
- ✓ Solute advection, adsorption and degradation
- ✓ One module \equiv one process or ensemble of processes on a landscape element
- ✓ Coupling of modules within the OpenPALM coupler (**Fouilloux1999**) turning the structure flexible



Twin experiments

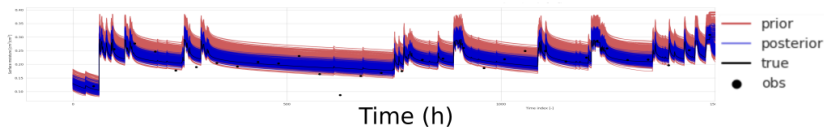
A virtual experiment where we know the true state : an output of the model

- $\text{obs } y = x^{\text{true}} + \text{error}$
- “data” = images of surface soil moisture \Rightarrow virtual data
- model = PESHMELBA
- errors are gaussian

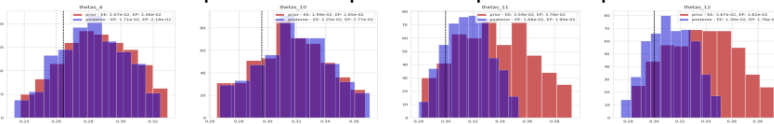


Twin experiments with PESHMELBA

Surface moisture - UH 480

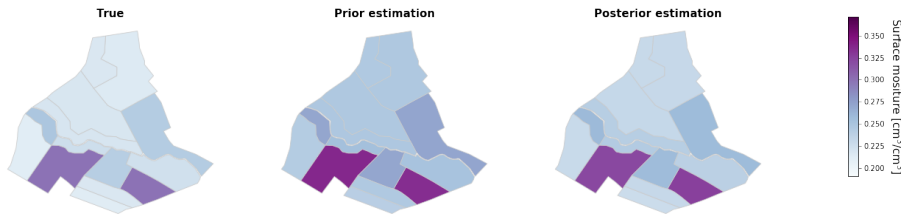


parameter prior and posterior pdf



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 \Rightarrow 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?)

\Rightarrow data assimilation = an optimal way to merge information

Thank you!
Any questions ?