Data assimilation to quantify and reduce uncertainty in ecohydrology modelling

Claire Lauvernet, Emilie Rouzies, Arthur Vidard, Alexandre Devers, Jean-Philippe Vidal, Laura Gatel, Katarina Radišić, Claudio Paniconi

Data assimilation to quantify and reduce uncertainty in ecohydrology modelling

séminaire ITES, 13 Nov. 2023

Claire Lauvernet[1], Emilie Rouzies[1], Arthur Vidard[2], Alexandre Devers[1], Jean-Philippe Vidal[1], Laura Gatel[2], Katarina Radišić[1], Claudio Paniconi[3]

[1]INRAE, UR RiverLy, Lyon
[2]INRIA, AIRSEA, Grenoble
[3]INRS/Univ. Laval, Québec, CA

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

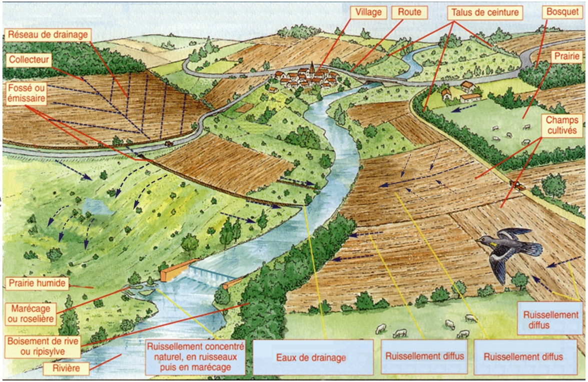
INRAE

Plan

Context : How to improve the water quality ?

⇒ a better understanding of water and pesticide transfer in soil

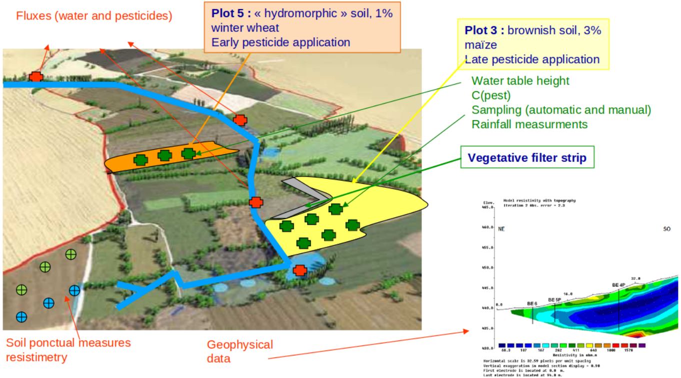
- Spatial heterogeneity of the soils, at all scales
- Soil and agricultural practices are more and more diverse
- Processes that drive the pesticide fate at the catchment scale are complex :
  - Hydrological transfer
  - adsorption
  - degradation



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

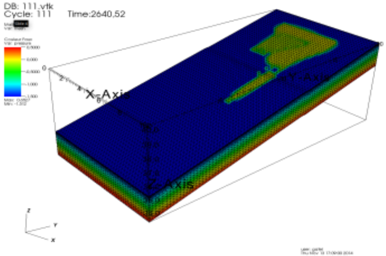


BUT without heavy experiments, this is very difficult to get the pesticides dynamics

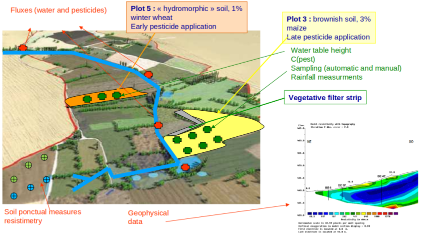
Spatially and temporally heterogeneous data. . .
. . . and pesticides modeling at several scales and several complexity degrees

- 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)

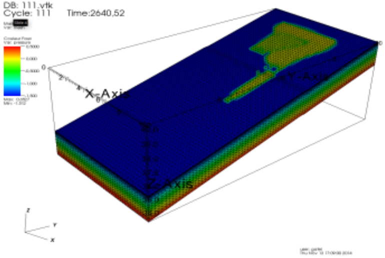
⇒ a high uncertainty (when we it is considered !)



Spatially heterogeneous data. . .



. . . and spatialized modeling



$\Rightarrow$ 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

Introduction: some challenges in ecohydrology

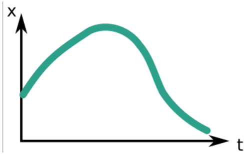
Data assimilation

An example of data assimilation in ecohydrological model

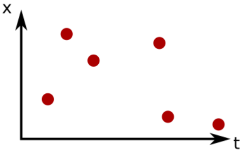
Data assimilation: definition

- 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)

Data assimilation: definition

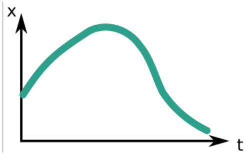


Model

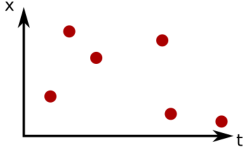


Observations

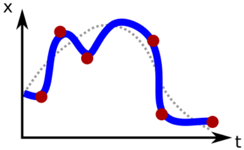
Data assimilation: definition



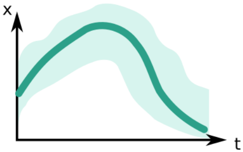
Model



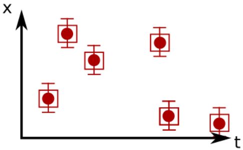
Observations



Data assimilation: definition

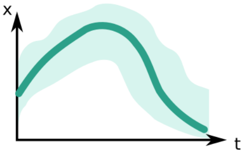


Model

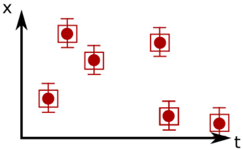


Observations

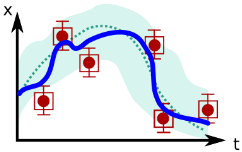
Data assimilation: definition



Model



Observations



Data assimilation: definition

"Approximation of the true state of a physical system at a given time by combining time-distributed observations with a dynamic model in an optimal way"(**Asch2016**)

⇒ **DA has two main goals**:

- optimally blend information from observations and model to produce an accurate and physically consistent estimate of the state of the system x^a
- quantify the uncertainty of this estimate for future users

Data assimilation: the ingredients

$\mathbf{x = (x_0, x_1, \ldots, x_N)^T}$ 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
- **the background** x^b is the state at $t_{k\text{-}1}$ and its associated error $\varepsilon^b = x^b - x^t_{k-1}$

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** $x_k = M_{k-1\rightarrow k}[x_{k-1}, param] + \eta_k$
- **the background** x^b is the state at t_{k-1} and its associated error $\varepsilon^b = x^b - x^t_{k-1}$
- **the observation model** $y_k = H_k[x_k] + \varepsilon_k$

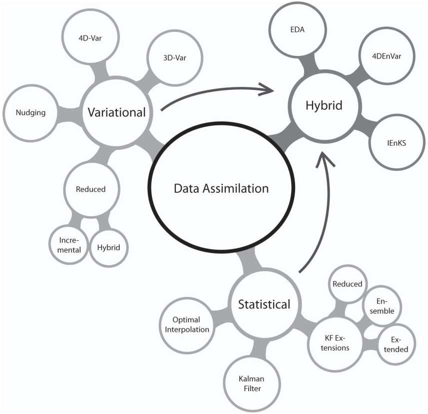
  y_k is the observation/data at time k
  ε_k the observation error, of covariance matrix R_k, e.g. instrumental error, representativeness
  $H : \mathcal{R}^m \rightarrow \mathcal{R}^d$ the observation operator that projects from model space to observational space (spatial interpolations, convolutions or spectral-to-physical space transformation in spectral models, etc.)

Data assimilation: the ingredients

$\mathbf{x} = (\mathrm{x_0,x_1,...,x_N})^T$ 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 background** x^b is the state at $t_{k\text{-}1}$ and its associated error $\varepsilon^b = x^b - x^t_{k-1}$
- **the observation model** $y_k = H_k[x_k] + \varepsilon_k$
- We assume that **model and obs. errors are random variables → described by pdf or by their covariance matrix**

Data assimilation: approaches



(from **Asch2016**)

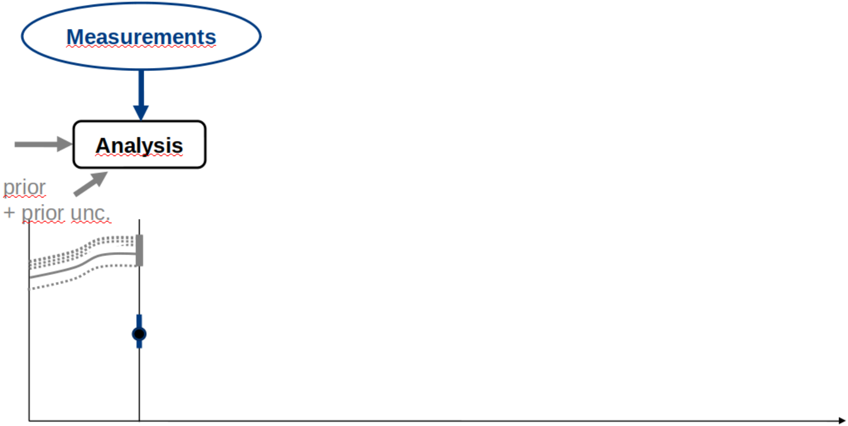
Data assimilation: approaches

- **Deterministic** methods (used in weather forecasting): state variables are assumed to be governed by deterministic laws.
  ⇒ methods based on optimal control, minimization of a cost function
- **Statistical/Stochastic** methods (used in ocean forecasting): a phenomenon is assumed to be the realization of a random variable: this is justified by the fact that the dynamics of the system under study (weather, ocean) are chaotic and therefore resemble a random system.
  ⇒ methods based on statistical estimation, Bayes theorem and Kalman filter
  ⇒ objective = determine a good approximation of the conditional expectation of the system state (as well as its error covariance matrix) given the observed data

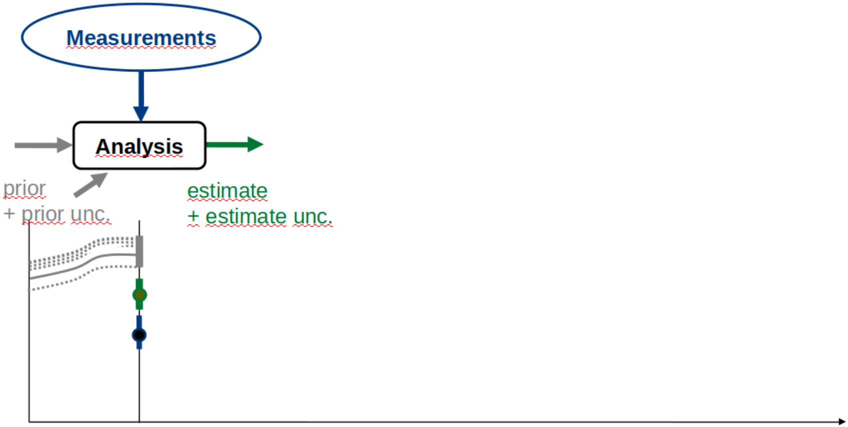
⇒ in a perfect context (linear, Gaussian, etc.), the methods are equivalent!

- In hydrology ? Chosen method should be suited to heterogeneous structure of the model, highly nonlinear processes but also suited to our high computational cost.

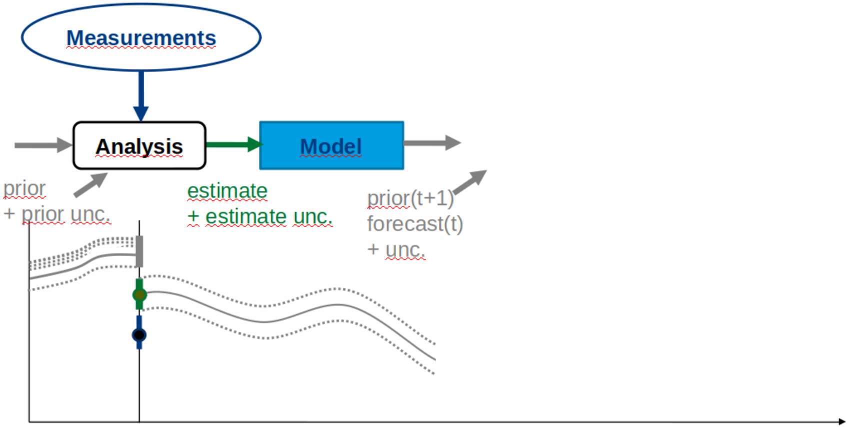
Data assimilation: definition



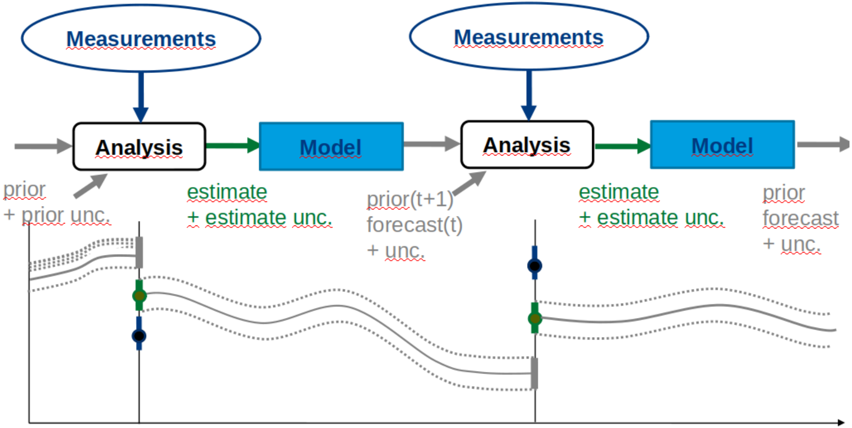
Data assimilation: definition



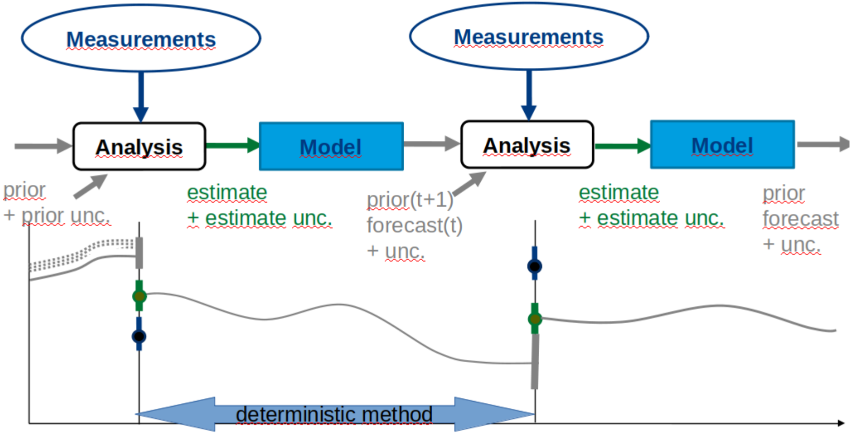
Data assimilation: definition



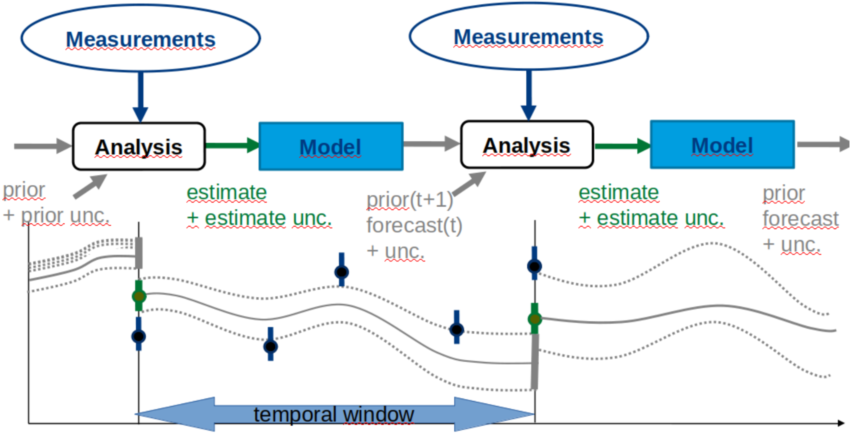
Data assimilation: definition



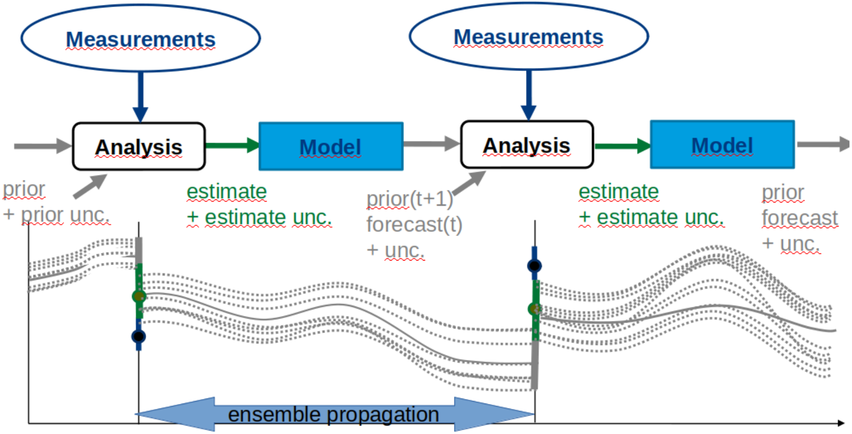
Data assimilation: definition



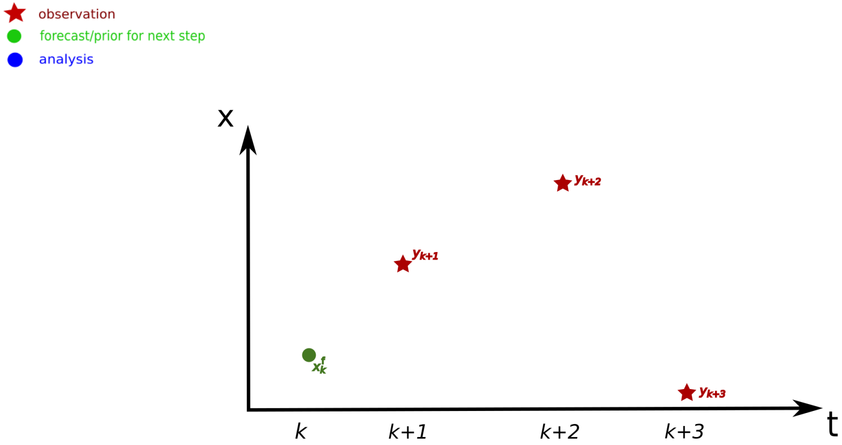
Data assimilation: definition



Data assimilation: definition

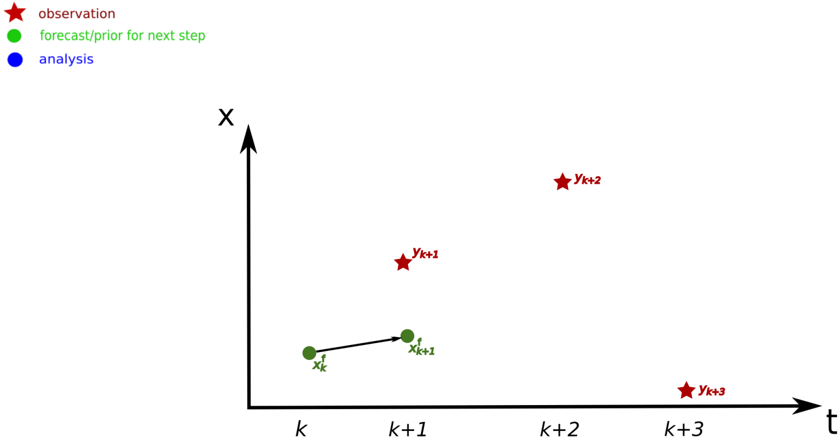


The Kalman Filter (**Kalman1960**) : estimate the optimal state at each observation time



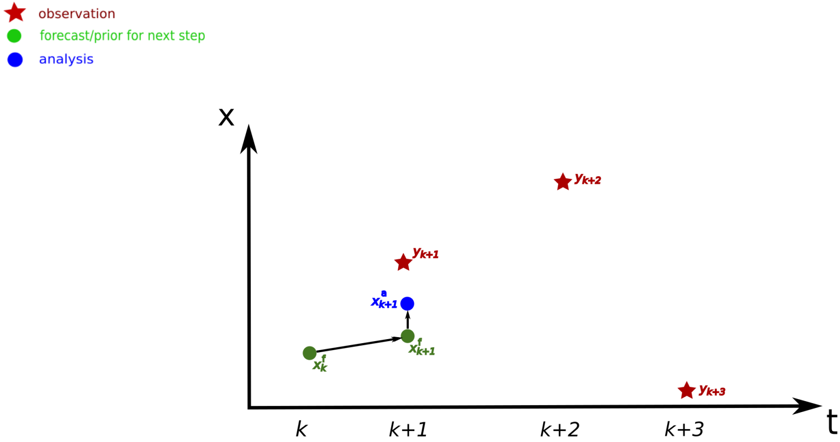
0. At time k: $\mathbf{X}_k^f$

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



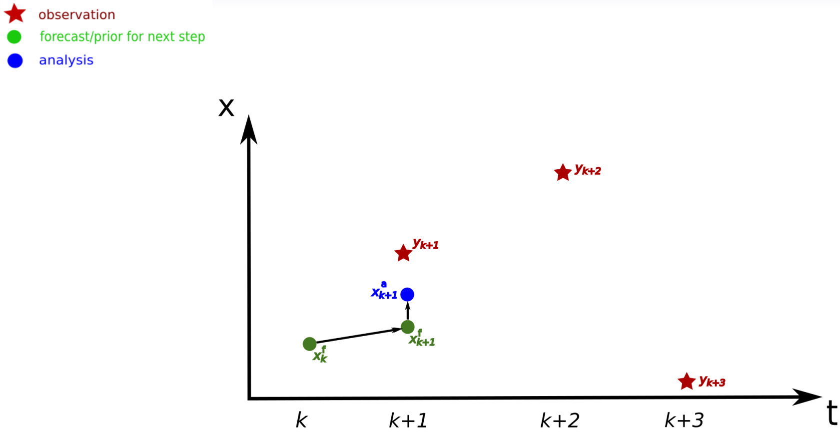
1. Forecast step: $\mathbf{X}_{k+1}^f = \mathbf{M}\mathbf{X}_k^f$

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



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

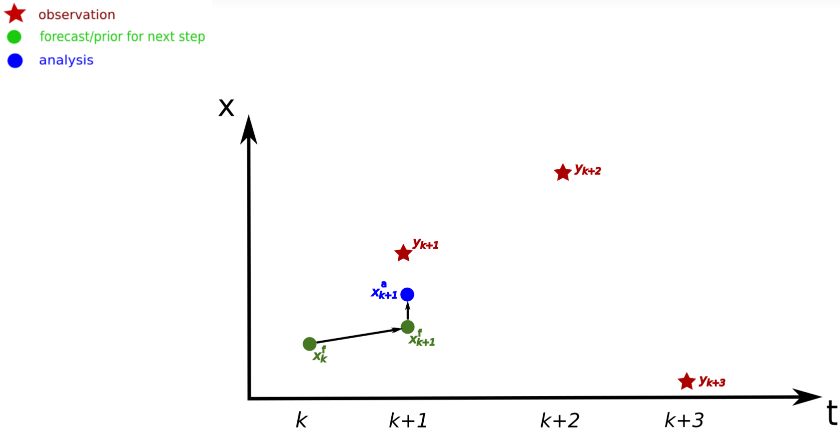
The Kalman Filter (**Kalman1960**) : estimate the optimal state at each observation time



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

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

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

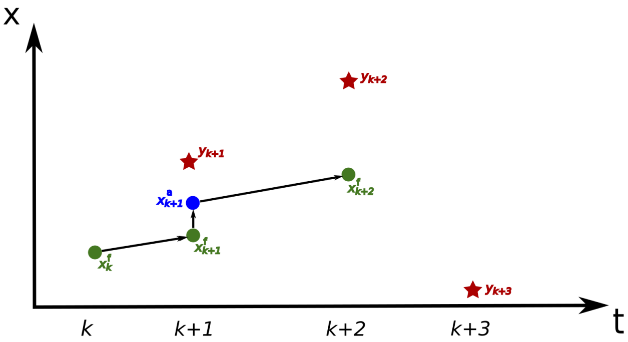


2. Analysis step: $\mathbf{X}^a_{k+1} = \mathbf{X}^f_{k+1} + \mathbf{K}_{k+1}(\mathbf{Y}_{k+1} - \mathbf{H}_{k+1}\mathbf{X}^f_{k+1})$

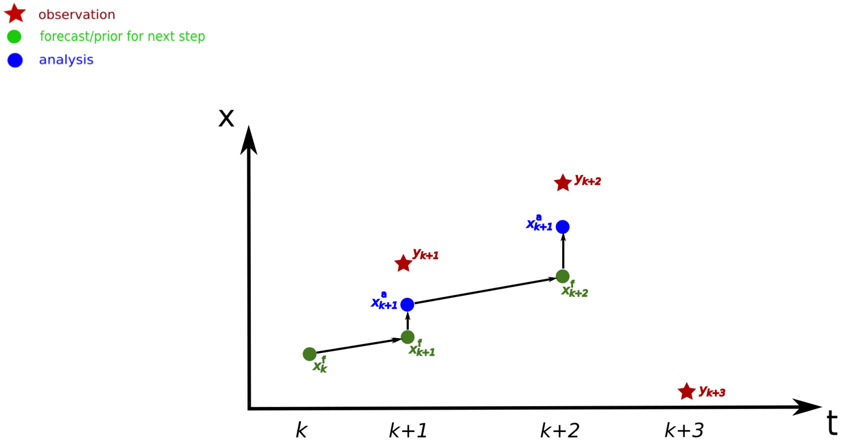
$$\mathbf{P}^a_{k+1} = \mathbf{P}^f_{k+1} - \mathbf{K}_{k+1}\mathbf{H}_{k+1}\mathbf{P}^f_{k+1}$$

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

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

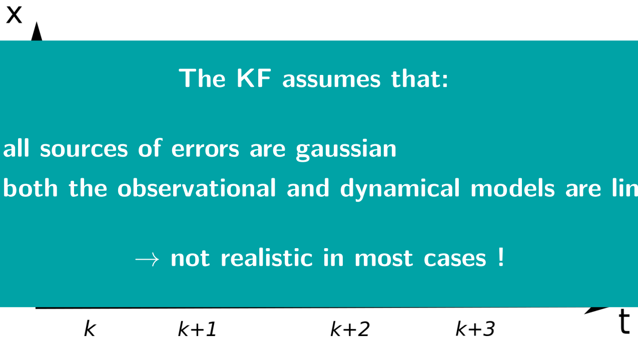


The Kalman Filter (**Kalman1960**) : estimate the optimal state at each observation time



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

- observation
- forecast/prior for next step
- analysis



The KF assumes that:

✓ **all sources of errors are gaussian**

✓ **both the observational and dynamical models are linear**

→ **not realistic in most cases !**

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

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

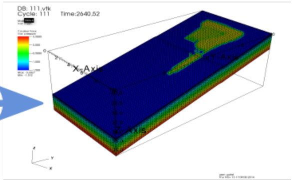
$$\frac{\partial Q}{\partial t} + c_k \frac{\partial Q}{\partial s} = D_h \frac{\partial^2 Q}{\partial s^2} + c_k q_s(h, \psi)$$

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

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

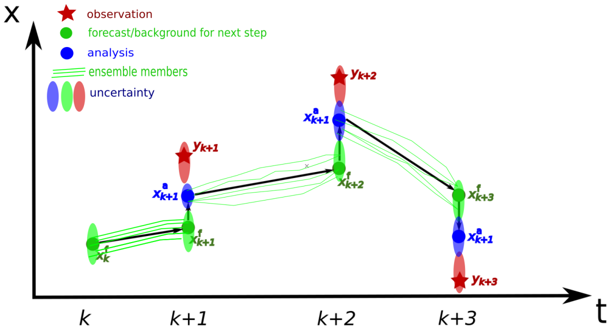
CATHY-Pesticides

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



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



On the variationnal side...

- Operational in meteorological centers : can deal with very large problems
- Use of ALL available information by solving a unique system
- Transform the inverse problem into an optimization problem The search of the minimum of the **cost function**

$$J = |sim. - obs.|^2 + apriori\ info$$

  is performed with standard methods (conjugate gradient, Newton-type methods)
- They require the computation of the gradient of J (optimality condition): $\nabla J = 0 \Leftrightarrow J$ is in an optimum

$\Rightarrow$ analytic approximation: computer time consuming

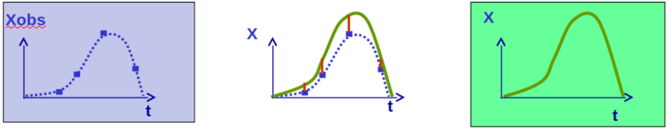
$\Rightarrow$ adjoint model method

. . . an optimization problem

model
State variable (streamflow, C°, soil moisture, ...)
parameters
Initial conditions
Cost function

$$\begin{cases} \dfrac{d\boxed{X}}{dt} & = \boxed{F}(\boxed{X}, \boxed{K}) \\ X(0) & = \boxed{u} \end{cases}$$

$$\boxed{J}(K) = \frac{1}{2}\|C.X(K) - X_{obs}\|^2_{\mathcal{X}_{obs}} + \frac{1}{2}\|K - K_0\|^2_{\mathcal{K}}$$



Problem : determine K^* that achieves the best LAI simulation closest to the observations, i.e. minimizing J

Variationnal data assimilation using the adjoint model

$$
\left\{
\begin{array}{rcl}
\boxed{\begin{array}{rcl} \dfrac{dX}{dt} & = & F(X,K) \\ X(0) & = & u \end{array}} & & \textbf{model} \\
\boxed{\begin{array}{rcl} \dfrac{dP}{dt} + \left[\dfrac{\partial F}{\partial X}\right]^T P & = & C^t(CX - X_{obs}) \\ P(T) & = & 0 \end{array}} & & \textbf{adjoint model} \\
\boxed{\begin{array}{rcl} \nabla_K J & = & -\left[\dfrac{\partial F}{\partial K}\right]^t P + K - K_0 \end{array}} & & \textbf{optimality condition}
\end{array}
\right.
$$

$\Rightarrow$ The optimality system contains all available information: observations, model, statistics ...

In practice, the gradient is computed by running an «adjoint model» derived from the model (automatic differentiation tool).

Plan

Introduction: some challenges in ecohydrology

Data assimilation

An example of data assimilation in ecohydrological model

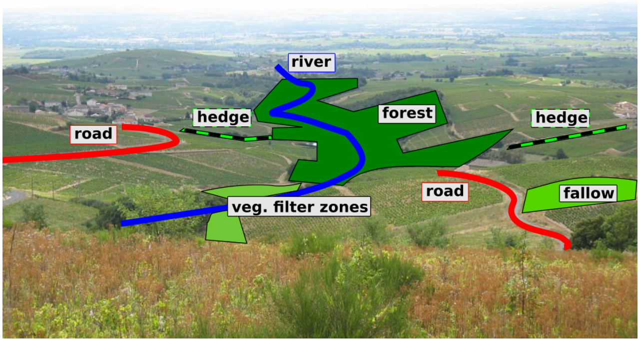
Context

Landscape features speed up or slow down pesticide transfer from the plots to the river.

⇒ The configuration of the catchment can influence the water quality.

Context

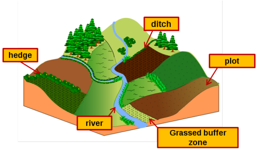
Landscape features speed up or slow down pesticide transfer from the plots to the river.



⇒ The configuration of the catchment can influence the water quality.

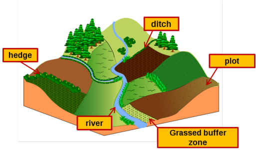
How to tackle pesticide transfers and fate on small agricultural catchments with modelling tools ?

✓ Integrating landscape elements diversity in a modular model

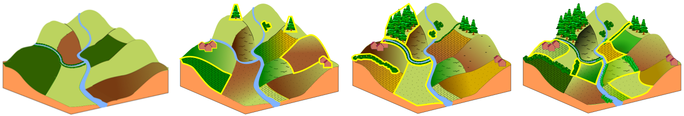


How to tackle pesticide transfers and fate on small agricultural catchments with modelling tools ?

✓ Integrating landscape elements diversity in a modular model

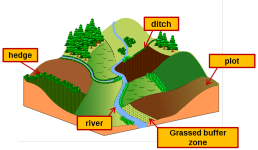


✓ Exploring landscape management scenarios

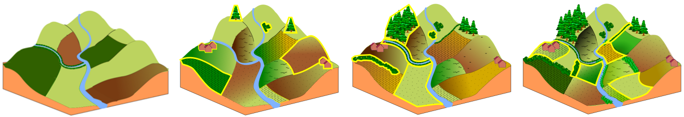


How to tackle pesticide transfers and fate on small agricultural catchments with modelling tools ?

✓ Integrating landscape elements diversity in a modular model



✓ Exploring landscape management scenarios

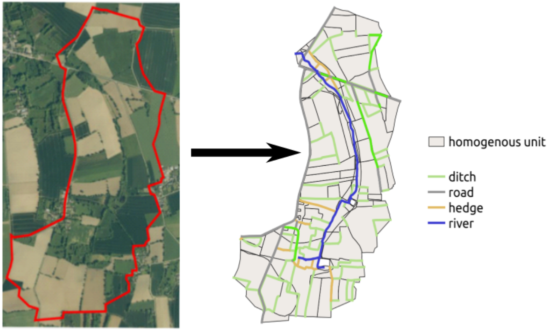


⇒ Development of the **PESHMELBA** model (**Rouzies2019**)

The PESHMELBA model (**Rouzies2019**)

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

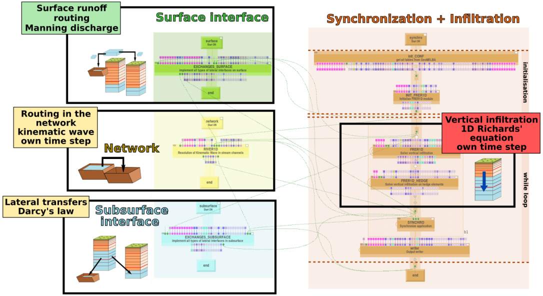
- ✓ Simulation of heterogenous landscapes composed of plots, vegetative filter zones, hedges, ditches and rivers
- ✓ Water transfers on surface and subsurface
- ✓ Solute advection, adsorption and degradation, solid transport



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, solid transport
- ✓ One module ≡ one process or ensemble of processes on a landscape element
- ✓ Coupling of modules within the OPENFLUID coupler to make it flexible



Uncertainty in PESHMELBA

We have a dream that one day PESHMELBA will be used as a decision-making tool to set up management scenarios and to identify an optimal landscape configuration for pesticide transfer mitigation.

Uncertainty in PESHMELBA

We have a dream that one day PESHMELBA will be used as a decision-making tool to set up management scenarios and to identify an optimal landscape configuration for pesticide transfer mitigation.

This is our objective...but before, it is necessary to **quantify** and **reduce** the uncertainty associated to PESHMELBA output variables.

Uncertainty in PESHMELBA

We have a dream that one day PESHMELBA will be used as a decision-making tool to set up management scenarios and to identify an optimal landscape configuration for pesticide transfer mitigation.

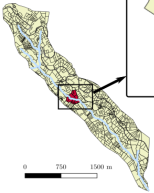
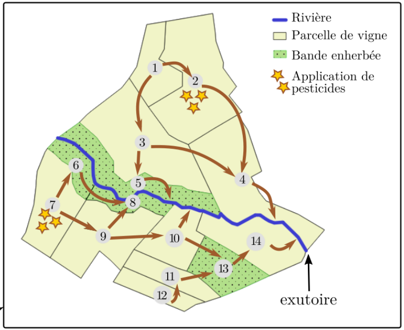
This is our objective...but before, it is necessary to **quantify** and **reduce** the uncertainty associated to PESHMELBA output variables.

Emilie Rouzies's PhD objectives

1. Quantify: performing an **uncertainty analysis** and a **sensitivity analysis** of the model
2. Reduce: performing **data assimilation** to integrate different sources of data: soil moisture images, ERT measurements and in-situ data of pesticide concentration

Case study

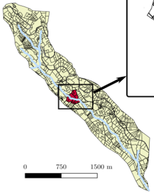
First attempt of DA in the PESHMELBA model: let's keep it simple...but realistic! (types of landscape elements, number of parameters, climate conditions...)

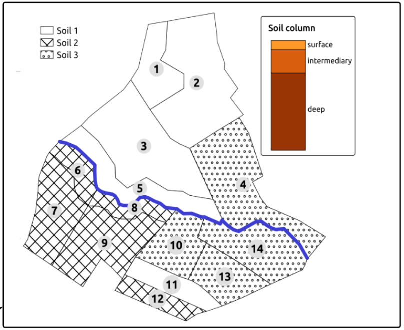



Panorama du bassin versant de la Morcille (Crédit photo: INRAE)

Case study

First attempt of DA in the PESHMELBA model: let’s keep it simple...but realistic! (types of landscape elements, number of parameters, climate conditions...)

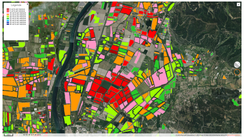




Panorama du bassin versant de la Morcille (Crédit photo: INRAE)

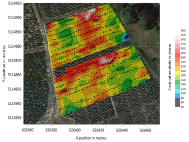
Which observations are available ?

Surface moisture images



- ✓ Surface moisture images got from the synergic use of Sentinel-1 and Sentinel-2 data
- ✓ One observation of mean moisture in the top 5 cm per landscape element per time step
- ✓ Freq. of observation: 144h, obs. error : assumed Gaussian, std $\sim 0.02\ cm^3cm^{-3}$ (! may highly differ on vineyard !)

In-situ moisture profiles



- ✓ Moisture profiles from EMI measurements or probe.
- ✓ Assumption : 2m-moisture profile on some landscape elements, obs. error : assumed Gaussian, std $\sim 0.02\ cm^3cm^{-3}$

Data assimilation

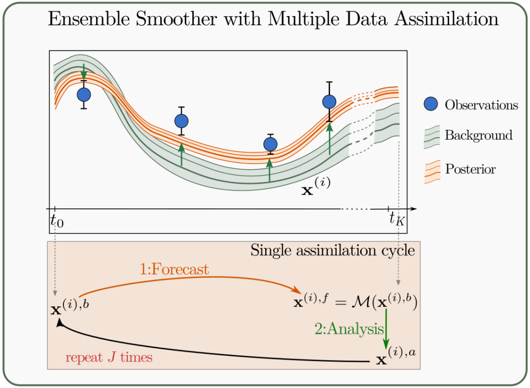
Objectives

- ✓ Improve moisture dynamics modelling **both in surface and subsurface**
- ✓ Improve estimation of pesticide export at the outlet
- ✓ Estimate input parameters (θ_{sat}) that would be set for the exploration of landscape management scenarios

DA method

Ensemble Smoother with Multiple Data Assimilation (**Emerick2013**)

- Ensemble method that inherits from Kalman Filter
- Iterative smoother well suited to parameter estimation problems in non linear contexts



Twin experiment

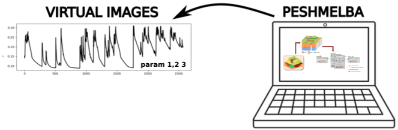
Twin experiment

First step: **twin experiments** to set and validate the DA framework (!! Reanalysis context)

Twin experiment

First step: **twin experiments** to set and validate the DA framework (!! Reanalysis context)

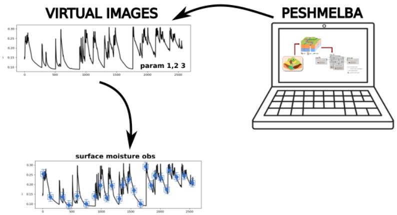
1. Use PESHMELBA to generate a "True" reference simulation



Twin experiment

First step: **twin experiments** to set and validate the DA framework (!! Reanalysis context)

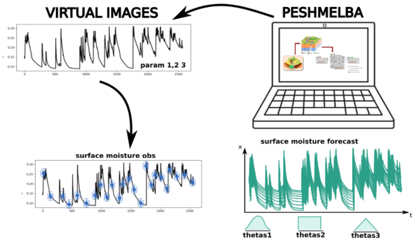
1. Use PESHMELBA to generate a "True" reference simulation
2. Add perturbation to generate virtual observations



Twin experiment

First step: **twin experiments** to set and validate the DA framework (!! Reanalysis context)

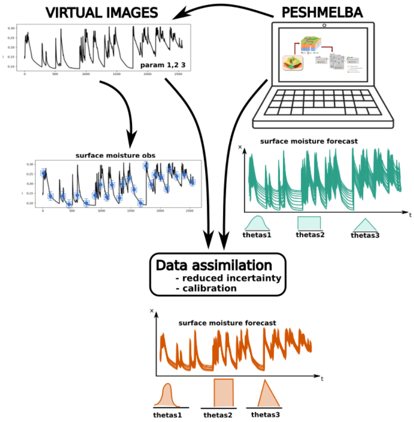
1. Use PESHMELBA to generate a "True" reference simulation
2. Add perturbation to generate virtual observations
3. Generate the prior/background state (build an ensemble from biased input parameters distribs)



Twin experiment

First step: **twin experiments** to set and validate the DA framework (!! Reanalysis context)

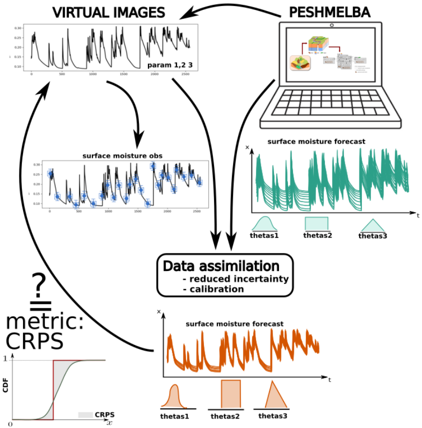
1. Use PESHMELBA to generate a "True" reference simulation
2. Add perturbation to generate virtual observations
3. Generate the prior/background state (build an ensemble from biased input parameters distribs)
4. Perform ensemble data assimilation to correct input parameters and moisture series towards the reference



Twin experiment

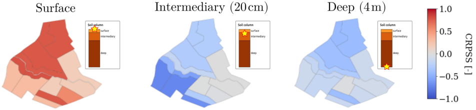
First step: **twin experiments** to set and validate the DA framework (!! Reanalysis context)

1. Use PESHMELBA to generate a "True" reference simulation
2. Add perturbation to generate virtual observations
3. Generate the prior/background state (build an ensemble from biased input parameters distribs)
4. Perform ensemble data assimilation to correct input parameters and moisture series towards the reference



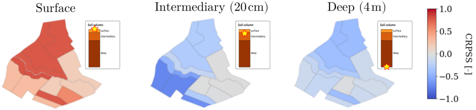
Results - Surface moisture images

- Moisture estimation

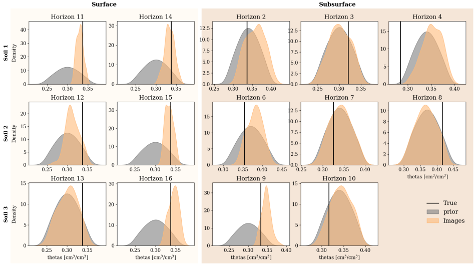


Results - Surface moisture images

- Moisture estimation

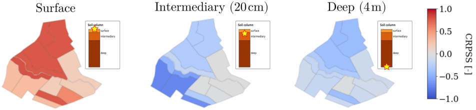


- Parameter estimation

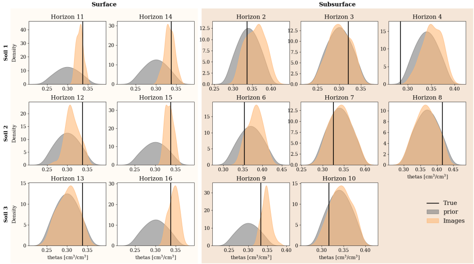


Results - Surface moisture images

- Moisture estimation



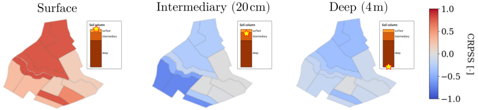
- Parameter estimation



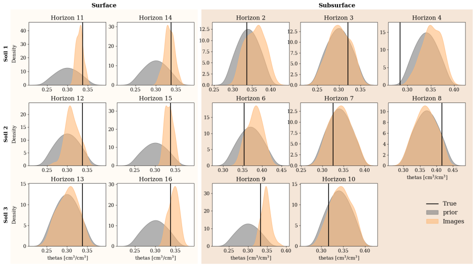
Data assimilation of satellite moisture images

Results - Surface moisture images

- Moisture estimation



- Parameter estimation



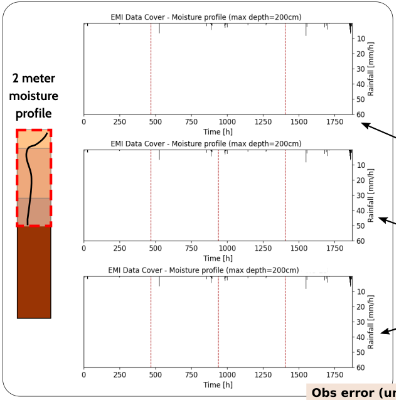
Data assimilation of satellite moisture images

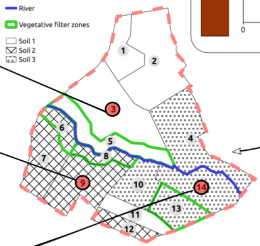
- Good correction of surface moisture and surface parameters
- Corrections do not propagate to subsurface (lack of correlations between surface and subsurface)

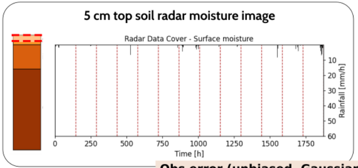
⇒ Idea? Integrate subsurface observations : point *vertical* profiles of moisture

INRAE

Results - Surface images + vertical profiles





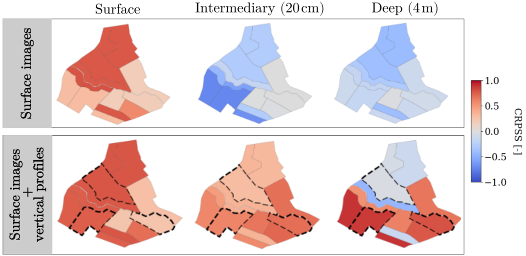


Obs error (unbiased, Gaussian)
std: 0.02cm^3/cm^3
Frequency: 6 days

Obs error (unbiased, Gaussian)
std: 0.02 cm^3/cm^3
Frequency: 2 or 3 ponctual profiles

Results - Surface images + vertical profiles

- Moisture estimation

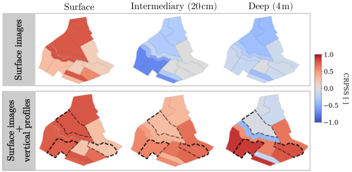


Adding **subsurface observations**

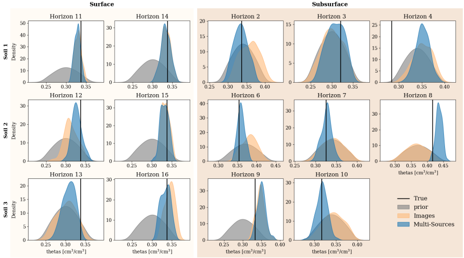
- improves moisture simulations at all depths

Results - Surface images + vertical profiles

- Moisture estimation



- Parameter estimation

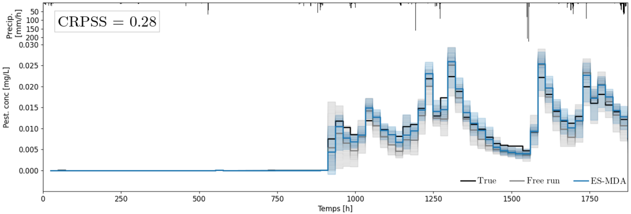


Adding **subsurface observations**

- improves moisture simulations at all depths
- improves θ_s estimates at all depths but on the plots of the same soil type

Impact of DA of hydological variables on pesticides variables ?

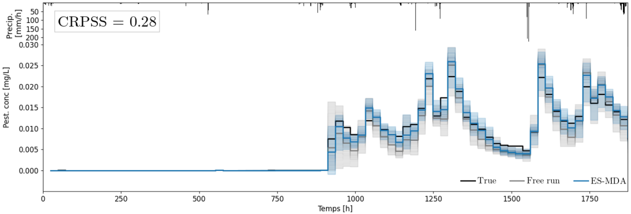
- Pesticide concentration at outlet



- (not shown) a relative impact of assimilating integrated concentration of pesticides (only if conc. are measured at high frequency ($<$ 5 days) and accurate)

Impact of DA of hydological variables on pesticides variables ?

- Pesticide concentration at outlet



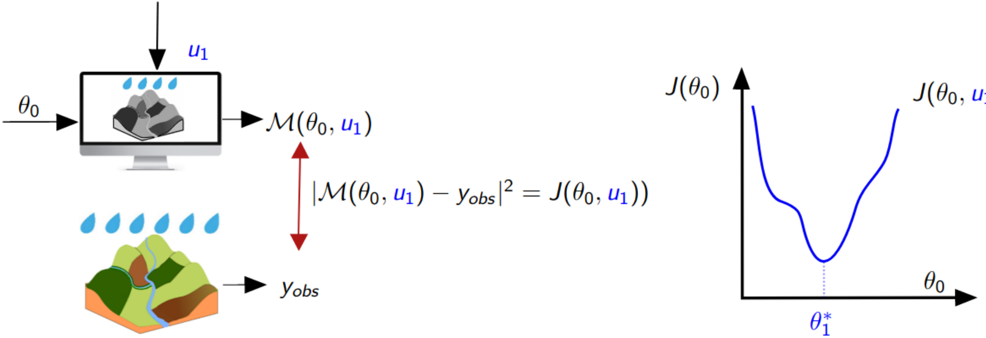
$\Rightarrow$ Coupled DA assimilation efficiently corrects pest. concentration.

- (not shown) a relative impact of assimilating integrated concentration of pesticides (only if conc. are measured at high frequency ($<$ 5 days) and accurate)

Conclusion

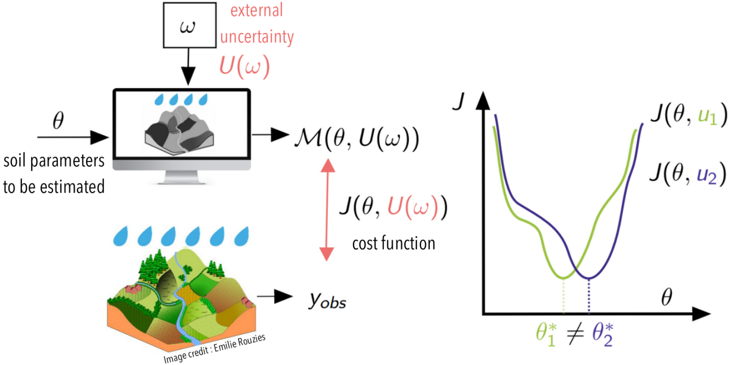
- DA framework set for the first time in PESHMELBA
- Multisource DA of hydrological data is efficient to also improve the pesticide transfer
- Next step : set a DA framework on a real catchment : many challenges ! (get data, characterize real observation errors, handle high computation cost...)
- Include external uncertainties such as forcings (rainfall, ETP, . . .)
- Compare with DA in the CATHY model, purely physics-based ($\Rightarrow$ less discontinuities?)

PhD Katarina Radišić : Calibrating a hydrological model robustly to rain perturbations (with stochastic surrogates)



What about the impact of external uncertainties on the calibration results?

PhD Katarina Radišić : Calibrating a hydrological model robustly to rain perturbations (with stochastic surrogates)



What does it mean to find a *robust* minimizer ?

References