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

To cite this version:
Claire Lauvernet, Emilie Rouzies, Arthur Vidard, Alexandre Devers, Jean-Philippe Vidal, et al.. Data assimilation to quantify and reduce uncertainty in ecohydrology modelling. Séminaire ITES 2023, ITES, Nov 2023, Strasbourg, France. hal-04387935

HAL Id: hal-04387935
https://hal.inrae.fr/hal-04387935
Submitted on 11 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License
Data assimilation to quantify and reduce uncertainty in ecohydrology modelling
séminaire ITES, 13 Nov. 2023

Claire Lauvernet¹, Emilie Rouzies¹, Arthur Vidard², Alexandre Devers¹, Jean-Philippe Vidal¹, Laura Gatel², Katarina Radišić¹, Claudio Paniconi³

¹INRAE, UR RiverLy, Lyon
²INRIA, AIRSEA, Grenoble
³INRS/Univ. Laval, Québec, CA
Introduction: some challenges in ecohydrology

Data assimilation

An example of data assimilation in ecohydrological model
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 measurements)
- 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... ⇒ 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
Introduction: some challenges in ecohydrology

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

\( x = (x_0, x_1, ..., x_N)^T \) 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** \( x_k = M_{k-1 \rightarrow k}[x_{k-1}, \text{param}] + \eta_k \)
  \( \eta_k \) the model error of covariance matrix \( P_k \)

- **the background** \( x^b \) is the state at \( t_{k-1} \) and its associated error \( \varepsilon^b = x^b - x_{k-1}^t \)
**Data assimilation: the ingredients**

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

- **the background** \( x^b \) is the state at \( t_{k-1} \) and its associated error \( \epsilon^b = x^b - x^t_{k-1} \)

- **the observation model** \( y_k = H_k[x_k] + \epsilon_k \)

\( y_k \) is the observation/data at time \( k \)

\( \epsilon_k \) 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

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

- **the background** \( x^b \) is the state at \( t_{k-1} \) and its associated error \( \varepsilon^b = x^b - x_{k-1}^t \)

- **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

Measurements → Analysis

prior + prior unc.
Data assimilation: definition

Measurements

Analysis

prior + prior unc.

estimate + estimate unc.
Data assimilation: definition

Data assimilation involves combining model predictions with observational data to improve the accuracy of the model. The process typically includes the following steps:

1. **Measurements** - Data collected from experiments or observations.
2. **Model** - A mathematical representation of the system being studied.
3. **Analysis** - The process of integrating the model and observations.

The assimilation process can be represented as:

- Prior state + Prior uncertainty → Estimate state + Estimate uncertainty → Prior state(t+1) + Forecast(t) + Uncertainty.
Data assimilation: definition

Measurements → Analysis

Model → Analysis

prior + prior unc
estimate + estimate unc
prior(t+1) forecast(t) + unc

estimate + estimate unc
prior forecast + unc
Data assimilation: definition

Measurements → Analysis → Model

Measurements → Analysis → Model

prior + prior unc. → estimate + estimate unc. → prior(t+1) forecast(t) + unc.

estimate + estimate unc. → prior forecast + unc.

Deterministic method
Data assimilation: definition
Data assimilation: definition

Measurements → Analysis → Model → Analysis → Model

prior + prior unc. → estimate + estimate unc. → prior(t+1) forecast(t) + unc. → estimate + estimate unc. → prior forecast + unc.

ensemble propagation
Introduction: some challenges in ecohydrology

Data assimilation

Application

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

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

0. At time $k$: $X^f_k$
The Kalman Filter (Kalman1960) : estimate the optimal state at each observation time

- Observation
- Forecast/prior for next step
- Analysis

1. Forecast step: $\mathbf{x}_{k+1}^f = \mathbf{Mx}_k^f$
The Kalman Filter (Kalman1960) : estimate the optimal state at each observation time

2. Analysis step: \( \mathbf{x}^a_{k+1} = \mathbf{x}^f_{k+1} + K_{k+1}(\mathbf{y}_{k+1} - H_{k+1}\mathbf{x}^f_{k+1}) \)
The Kalman Filter (Kalman1960): estimate the optimal state at each observation time

2. Analysis step:

\[ X_{k+1}^a = X_{k+1}^f + K_{k+1}(Y_{k+1} - H_{k+1}X_{k+1}^f) \]

with

\[ K_{k+1} = P_{k+1}^f H_{k+1}^T [H_{k+1} P_{k+1}^f H_{k+1}^T + R_{k+1}]^{-1} \]
The Kalman Filter ([Kalman1960]) : estimate the optimal state at each observation time

2. Analysis step: $X_{k+1}^a = X_{k+1}^f + K_{k+1}(Y_{k+1} - H_{k+1}X_{k+1}^f)$

$$P_{k+1}^a = P_{k+1}^f - K_{k+1}H_{k+1}P_{k+1}^f$$
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

- observation
- forecast/prior for next step
- analysis
The Kalman Filter (Kalman 1960): 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 $\Rightarrow$ discontinuities, thresholds
  $\rightarrow$ **definitely not gaussian**!
  $\rightarrow$ **Ensemble filter approaches**

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**
  \[ J = |\text{sim.} - \text{obs.}|^2 + \text{apriori info} \]
  is performed with standard methods (conjugate gradient, Newton-type methods)
- They require the computation of the gradient of J (optimality condition):
  \[ \nabla J = 0 \iff J \text{ is in an optimum} \]
  \[ \implies \text{analytic approximation: computer time consuming} \]
  \[ \implies \text{adjoint model method} \]
... an optimization problem

\[
\begin{align*}
\frac{dX}{dt} &= F(X, K) \\
X(0) &= u
\end{align*}
\]

Cost function

\[
J(K) = \frac{1}{2} \| C X(K) - X_{obs} \|_{X_{obs}}^2 + \frac{1}{2} \| K - K_0 \|_K^2
\]

Problem: determine \( K^* \) that achieves the best LAI simulation closest to the observations, i.e. minimizing \( J \)
Variational data assimilation using the adjoint model

\[
\begin{align*}
\frac{dX}{dt} &= F(X, K) \\
X(0) &= u \\
\frac{dP}{dt} + \left[ \frac{\partial F}{\partial X} \right]^T P &= C^t (CX - X_{obs}) \\
P(T) &= 0 \\
\nabla_K J &= - \left[ \frac{\partial F}{\partial K} \right]^t P + K - K_0
\end{align*}
\]

⇒ 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).
Introduction: some challenges in ecohydrology

Data assimilation

An example of data assimilation in ecohydrological model
Introduction: some challenges in ecohydrology

Data assimilation

Application

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.
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

![Diagram showing various landscape elements such as hedges, plots, rivers, and grassed buffer zones.](image-url)
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

⇒ Development of the PESHMELBA model (Rouzies2019)

✓ Exploring landscape management scenarios
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 heterogeneous 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...)
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 $\sim 0.02 \text{ cm}^3 \text{ cm}^{-3}$ (! 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 $\sim 0.02 \text{ cm}^3 \text{ cm}^{-3}$
Data assimilation

Objectives

✓ Improve moisture dynamics modelling both in surface and subsurface
✓ Improve estimation of pesticide export at the outlet
✓ Estimate input parameters ($\theta_{sat}$) that would be set for the exploration of landscape management scenarios
Introduction: some challenges in ecohydrology

Data assimilation

Application

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

Ensemble Smoother with Multiple Data Assimilation

- Observations
- Background
- Posterior

Single assimilation cycle

1: Forecast
2: Analysis

repeat J times
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
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 distris)
### 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 distris)
4. Perform ensemble data assimilation to correct input parameters and moisture series towards the reference
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

![Surface](image1)

![Intermediary (20cm)](image2)

![Deep (4m)](image3)

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.
Introduction: some challenges in ecohydrology

Data assimilation

Application

Results - Surface moisture images

- Moisture estimation
  - Surface
  - Intermediary (20cm)
  - Deep (4m)

- Parameter estimation

  ![Surface Moisture Estimation](Diagram1)

  ![Subsurface Parameter Estimation](Diagram2)

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.
Introduction: some challenges in ecohydrology

Data assimilation

Application

Results - Surface moisture images

- Moisture estimation
  - Surface
  - Intermediary (20cm)
  - Deep (4m)

Data assimilation of satellite moisture images

- Parameter estimation
Introduction: some challenges in ecohydrology

Data assimilation

Application

Results - Surface moisture images

• Moisture estimation
  - Surface
  - Intermediary (20 cm)
  - Deep (4 m)

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
Introduction: some challenges in ecohydrology

Data assimilation

Application

Results - Surface images + vertical profiles

Obs error (unbiased, Gaussian)

std: 0.02 cm$^3$/cm$^3$

Frequency: 6 days

Obs error (unbiased, Gaussian)

std: 0.02 cm$^3$/cm$^3$

Frequency: 2 or 3 punctual profiles
Results - Surface images + vertical profiles

• Moisture estimation

Adding **subsurface observations**

• improves moisture simulations at all depths
Introduction: some challenges in ecohydrology

Data assimilation

Application

Results - Surface images + vertical profiles

- Moisture estimation

- Parameter estimation

Adding subsurface observations

- improves moisture simulations at all depths
- improves $\theta_s$ 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

 CRPSS = 0.28

- (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 hydrological variables on pesticides variables?

- Pesticide concentration at outlet

⇒ 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 (⇒ 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?