Data assimilation in glaciology

Bertrand Bonan[1] Maëlle Nodet[1] Catherine Ritz[2]

[1]Université de Grenoble, INRIA, LJK

[2]Université de Grenoble, CNRS, LGGE

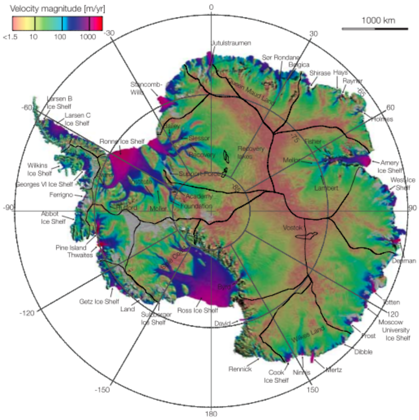
Journée de rencontres OSUG – LJK
18 mars 2013

Sea level change: Antarctica & Greenland contribution

Motivations: Computation of the ice discharge of Antarctica and Greenland in the near future, thanks to simulations of polar ice sheet model.

Ice discharge:

- governed by a couple of narrow outlets (ice streams),
- closely linked to ice velocities,
- highly sensitive to basal friction parameters,
- highly sensitive to bedrock topography



Surface ice velocities [Rignot *et al.* 2011] →

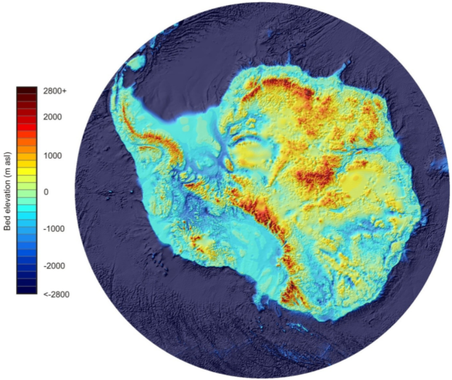
Poorly known basal parameters

Basal drag and bedrock topography are crucial to perform accurate simulations of ice sheets.

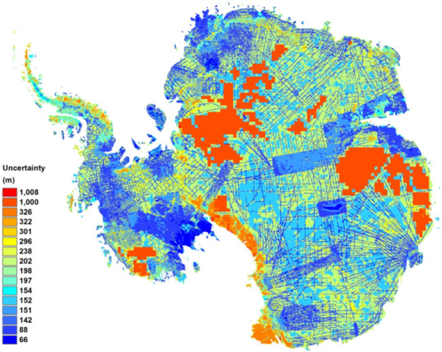
But:

- basal drag is unknown (local estimation by analysis of extracted sediment, poorly representative lab experiments, geothermal flux impacting basal temperature not well known)
- bedrock topography is measured along tracks $\Longrightarrow$ up to 400-500 meters uncertainties on central regions of Greenland / the Antarctica

Example: Bedmap2 [Fretwell *et al.* 2012]

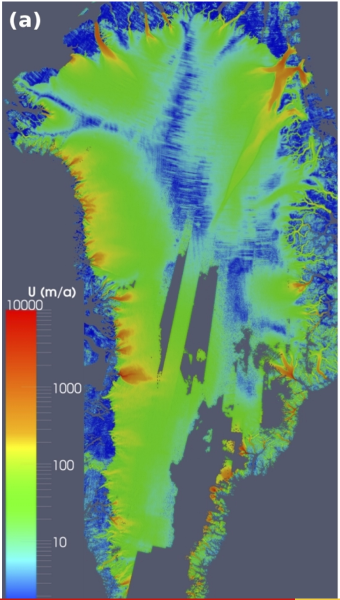


Bedmap2 estimation of bed topography



Estimated uncertainty of bed topography

Data assimilation



Combine model equations and observations:

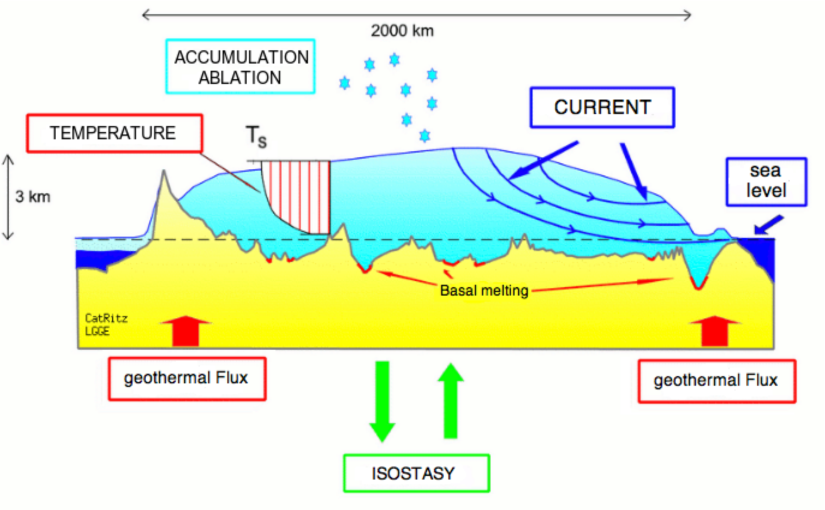
- surface velocities,
- surface elevation,
- surface trends,
- bedrock topography.

in order to infer basal drag and bedrock topography

Outline

Where am I?

Ice dynamics processes



Model equations: mass balance

Large time and space scales ⇒ shallow model

Example: flowline SIA+SSA model (1D + time)

Mass balance equation:

$$\frac{\partial H}{\partial t} = \dot{b}_m - \frac{\partial \left(\overline{U} H\right)}{\partial x}, \quad H|_{t=0} = H_0, \quad H \geq 0$$

with

- x latitude, t time
- $H(t,x)$ ice thickness, $H_0(x)$ initial ice thickness
- $\overline{U}(t,x)$ ice velocity averaged over ice thickness:

$$\overline{U}(t,x) = \frac{1}{H(t,x)} \int_{B(t,x)}^{H(t,x)} u(t,z)\, dz$$

  where $B(t,x)$ is the bedrock topography
- $\dot{b}_m(t,x)$ surface mass balance rate

Model equations: dynamics (1)

Vertically averaged ice velocity is a diagnostic variable → no partial derivative in time involved, computed from geometry at each time step

$$\overline{U} = U_d + U_s$$

U_d deformation contribution, U_s sliding contribution.

→ Deformation contribution:

$$U_d = -a_1 \frac{\partial S}{\partial x} \frac{H^2}{3} - a_2 \left(\frac{\partial S}{\partial x} \right)^3 \frac{H^4}{3}, \quad S = B + H$$

with

- $S(t,x)$ ice surface elevation, $B(t,x)$ ice bottom elevation ;
- a_1, a_2 coefficients (may vary).

Model equations: dynamics (2)

$\rightarrow$ Sliding contribution:

- In case of ice-streams or ice-shelves, U_s solution of

$$\frac{\partial}{\partial x}\left(4H\eta\frac{\partial U_s}{\partial x}\right) = \rho g H \frac{\partial S}{\partial x} - \tau_b$$

  with:

  - η effective ice viscosity
  - $\tau_b = -\beta U_s$ basal shear stress
  - $\beta > 0$ basal friction coefficient

- else, $U_s = 0$

β models bedrock properties: sediment, rock, rock debris, cold or melting ice, presence of thin water layer, subglacial lake or river, water cavities...

Where am I?

What is data assimilation?

Combine at best different sources of information to estimate the state of a system:

- model equations
- observations, data
- background, a priori information
- statistics

What is data assimilation for?

Historically: initial state estimation, for weather forecasting.

Today, many other applications:

- initial conditions for predictions,
- calibration and validation,
- observing system design, monitoring and assessment,
- reanalysis,
- better understanding (model errors, data errors, physical process interactions, parameters, etc),
- etc.

And many other fields:

- oceanography
- glaciology,
- seismology,
- nuclear fusion,
- medicine,
- agronomy,
- etc.

Framework of DA: least squares analysis

Aim: solve the inverse problem $\mathbf{y}^o = \mathcal{H}(\mathbf{x}^t) + \epsilon^o$, given a background estimate $\mathbf{x}^f$ of the true input parameters $\mathbf{x}^t$, where:

- $\mathbf{y}^o$ are incomplete observations, with errors ϵ^o unbiased and non trivial, with covariance matrix $\mathbf{R}$ given.
- $\mathbf{x}^f = \mathbf{x}^t + \epsilon^f$, ϵ^f background errors unbiased and non trivial, with covariance matrix $\mathbf{P}^f$ given
- observation operator $\mathcal{H}$ maps the input parameters to the observation variables (can contain complex laws, PDEs, non linear physics, ...)

Hyp: $\mathcal{H} = \mathbf{H}$ is a linear operator, ϵ^o and ϵ^f are not correlated.

$\rightarrow$ The estimate $\mathbf{x}^a$ of $\mathbf{x}^t$ is searched for as a linear combination:

$$\mathbf{x}^a = \mathbf{L}\,\mathbf{x}^f + \mathbf{K}\,\mathbf{y}^o$$

with the *optimality criterium*: unbiased estimate $\mathbf{x}^a$, with minimal variance $\mathrm{tr}(\mathbf{P}^a)$.

Best linear unbiased estimator, or least squares analysis

1. BLUE analysis:

$$\begin{cases} \mathbf{x}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\,\mathbf{x}^f + \mathbf{K}\,\mathbf{y}^o = \mathbf{x}^f + \mathbf{K}(\mathbf{y}^o - \mathbf{H}(\mathbf{x}^f)) \\ \mathbf{K} = \mathbf{P}^f\mathbf{H}^T(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R})^{-1} \end{cases}$$

   $\mathbf{K}$: gain, or weight matrix, $\mathbf{y}^o - \mathbf{H}(\mathbf{x}^f)$ innovation.

2. Analysis covariance matrix: $\mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^f$

3. Equivalent variational optimization problem: (optimal least squares)

$$\begin{cases} \mathbf{x}^a = \arg\min \mathcal{J} \\ \mathcal{J}(\mathbf{x}) = (\mathbf{x} - \mathbf{x}^f)^T\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^f) + (\mathbf{y}^o - \mathbf{H}(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{y}^o - \mathbf{H}(\mathbf{x})) \end{cases}$$

   $\mathcal{J}$: cost function, inverse of Hessian of $\mathcal{J}$ at $\mathbf{x}^a$: $\mathbf{P}^a$

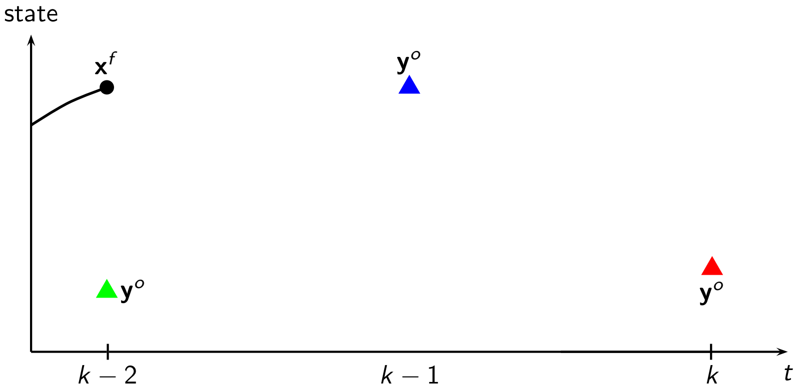
Data assimilation methods

Two types of methods:

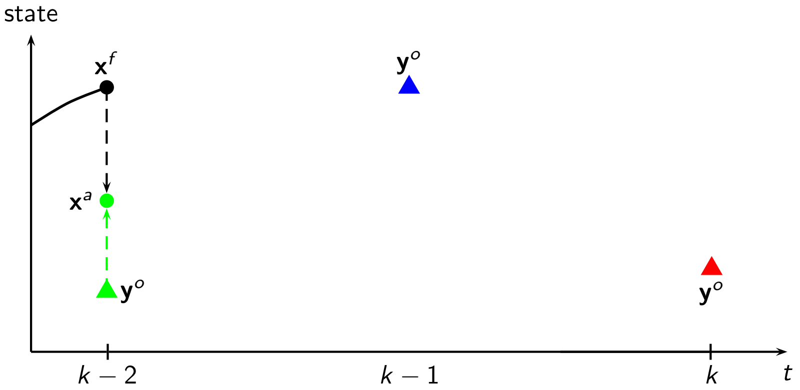
1. Direct computation of the BLUE, and the gain matrix $\mathbf{K}$.
   Main algorithm: Kalman filter
   $\longrightarrow$ stochastic data assimilation.

2. Minimization of the cost function $\mathcal{J}$ using optimization and adjoint methods.
   Main algorithm: 4D-Var
   $\longrightarrow$ variational data assimilation.

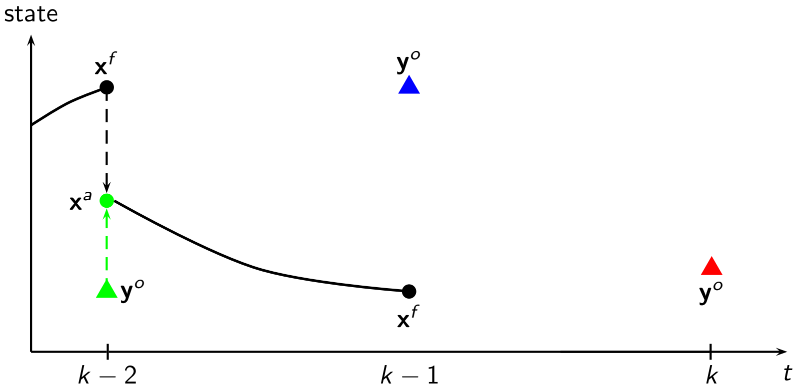
Sequential data assimilation: Kalman filter sequence



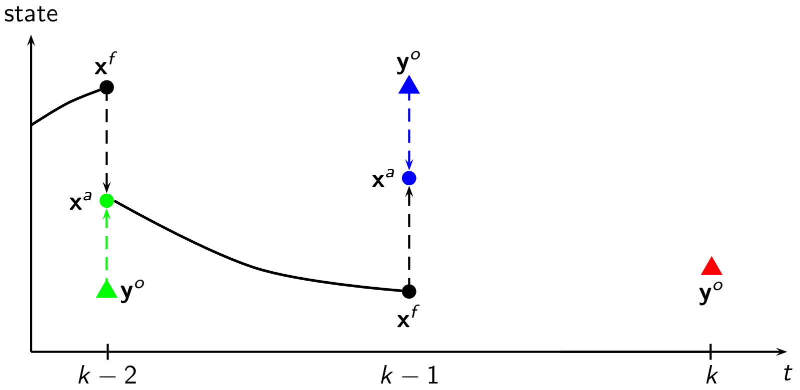
Sequential data assimilation: Kalman filter sequence



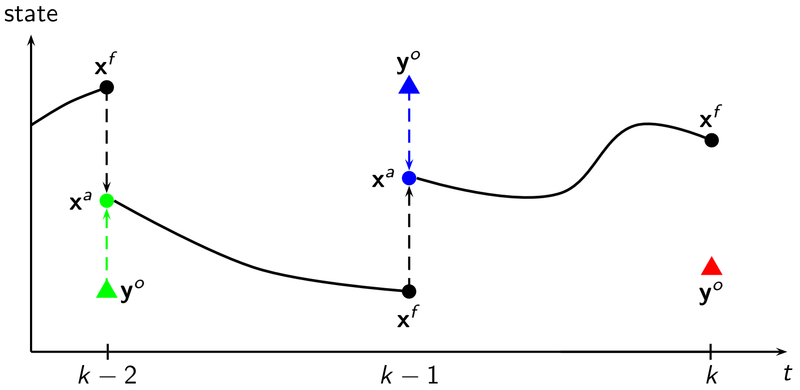
Sequential data assimilation: Kalman filter sequence



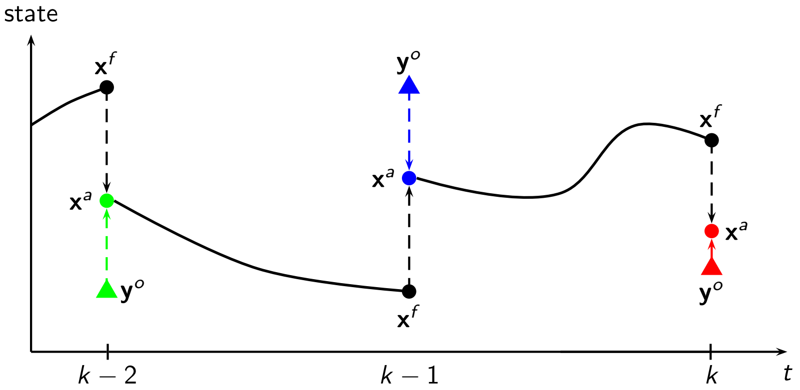
Sequential data assimilation: Kalman filter sequence



Sequential data assimilation: Kalman filter sequence



Sequential data assimilation: Kalman filter sequence



Kalman filter

State vector pdf represented by 2 variables:

- $\mathbf{x}_k$ state estimate at time t_k
  (in our case, bedrock topography and ice thickness)
- $\mathbf{P}_k$ error covariance matrix
  (a measure of the estimated accuracy of the state estimate)

Two phases

- **Forecast**: [Model free run]
  Use $(\mathbf{x}^a_{k-1}, \mathbf{P}^a_{k-1})$ to produce an estimation at current time t_k thanks to model $\mathcal{M}_k$. We obtain $(\mathbf{x}^f_k, \mathbf{P}^f_k)$.
- **Analysis**: [BLUE]
  Update $(\mathbf{x}^f_k, \mathbf{P}^f_k)$ with observations $\mathbf{y}^o_k$, error covariance matrix $\mathbf{R}_k$ and observation operator $\mathcal{H}_k$. We obtain $(\mathbf{x}^a_k, \mathbf{P}^a_k)$.

Initialisation with a priori information
+ Several assumptions needed for optimality ▸ details

Ensemble Kalman Filter

KF unpractical for geophysical data assimilation: **P** matrices too large to be computed/stored

$\Rightarrow$ Use Monte-Carlo method: Ensemble Kalman Filter (EnKF) [Evensen 1994]

A small set of state vectors representative of the model $\left\{\mathbf{x}^{(i)}, i = 1, \ldots, N_{ens}\right\}$ is used to approximate model mean and covariances:

$$\begin{aligned}
\mathbf{x} &\approx \bar{\mathbf{x}} = \frac{1}{N_{ens}} \sum_{i=1}^{N_{ens}} \mathbf{x}^{(i)} \\
\mathbf{P} &\approx \mathbf{P}_e = \frac{1}{N_{ens} - 1} \sum_{i=1}^{N_{ens}} \left(\mathbf{x}^{(i)} - \bar{\mathbf{x}}\right) \left(\mathbf{x}^{(i)} - \bar{\mathbf{x}}\right)^T
\end{aligned}$$

A small ensemble ($N_{ens} = O(100)$) give usually satisfying results.

Here we use a deterministic EnKF called ETKF.

more details on ETKF

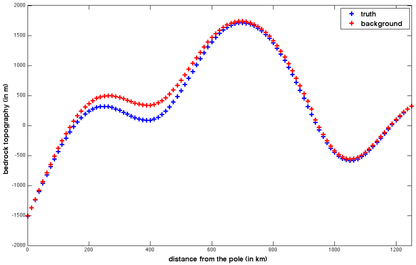
Where am I?

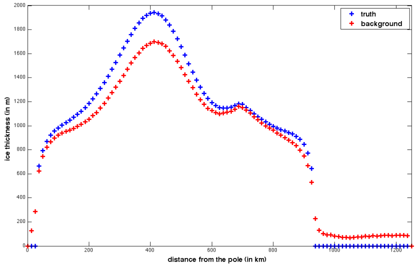
What we want: control variables

We search the actual state of our model governed entirely by B and H

Twin experiments

Simulate data thanks to the model and do data assimilation to retrieve the state variables and/or parameters you choose to simulate data.

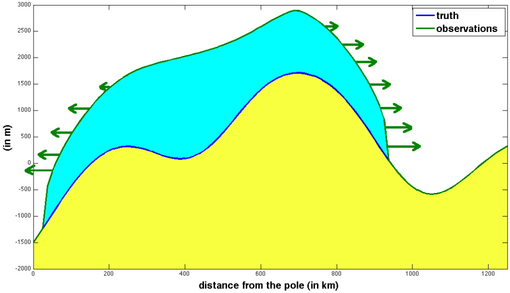




What we have: surface observations

We observe each year during 10 years

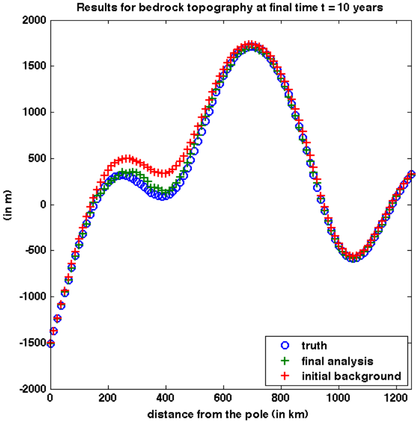
- surface elevation ($\sigma_S = 10$ m)
- surface velocity ($\sigma_u = 3$ m/yr)

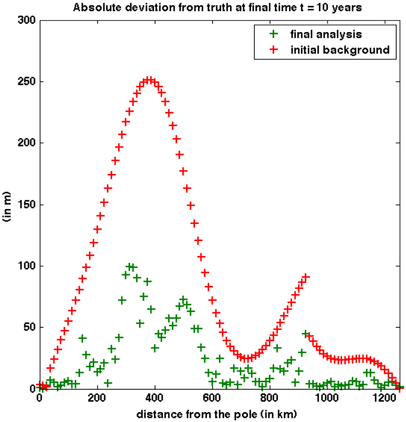


Obs. taken at each grid point except in the centre every 2 grid points.

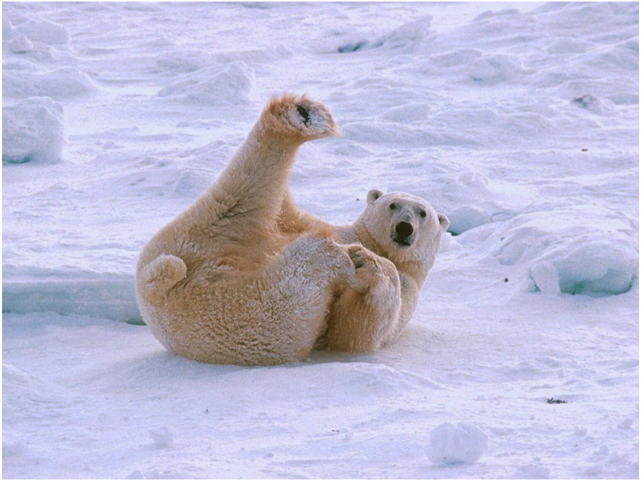
Noisy observations (noise in compliance with σ_S and σ_u)

LETKF with inflation and localisation $N_{ens} = 10$





Thank you for your attention



Where am I?

Where are we?

4 Appendix
- How to build initial ensemble?
- Kalman filter details
- ETKF details

How to build initial ensemble?

In our case, well-known surface elevation $S = B + H$
But poor information on B and H. We use these two facts.

Assume B^{back} a priori information on bedrock topography.

For member (i), we compute $[B^{(i)}, H^{(i)}]$:

- $B^{(i)} = B^{back} + \mathbf{b}^{(i)}$ with $\mathbf{b}^{(i)} \sim \mathcal{N}(0, \mathbf{Cov}_B)$ with a good length scale for space correlation.
- $S^{(i)} = S^{obs} + \mathbf{s}^{(i)}$ with $\mathbf{s}^{(i)} \sim \mathcal{N}(0, \sigma_S^2 \mathbf{I})$.
- $H^{(i)} = S^{(i)} - B^{(i)}$ then run the model for 10 years and take resulting ice thickness for $H^{(i)}$ (which is more physical).

Where are we?

KF and EKF hypothesis

- Initial state is gaussian $\sim \mathcal{N}(\mathbf{x}^b, \mathbf{B})$.
- Dynamical model $\mathcal{M}_k$ is linear (KF) or can be linearized (EKF) with its tangent linear matrix $\mathbf{M}_k$ (EKF).
- Model errors are unbiased and gaussian $\sim \mathcal{N}(0, \mathbf{Q}_k)$.
- Model errors are uncorrelated in time.
- Observation operators $\mathcal{H}_k$ are linear (KF) or can be linearized with its tangent linear matrix $\mathbf{H}_k$ (EKF).
- Observation errors are unbiased and gaussian $\sim \mathcal{N}(0, \mathbf{R}_k)$.
- Observations errors are uncorrelated in time.
- Errors of different types are independant.

Algorithm

1. Initialisation: $\mathbf{x}_0^f = \mathbf{x}^b$ and $\mathbf{P}^f = \mathbf{B}$.
2. Forecast step: (evolution model)

$$\begin{aligned} \mathbf{x}_{k+1}^f &= \mathcal{M}_k\left(\mathbf{x}_k^a\right) \\ \mathbf{P}_{k+1}^f &= \mathbf{M}_k\,\mathbf{P}_k^a\,\mathbf{M}_k{}^T + \mathbf{Q}_k \end{aligned}$$

3. Analysis step: (BLUE analysis)

$$\begin{aligned} \mathbf{x}_k^a &= \mathbf{x}_k^f + \mathbf{K}_k\left(\mathbf{y}_k^o - \mathcal{H}_k\left(\mathbf{x}_k^f\right)\right) \\ \mathbf{P}_k^a &= \left(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k\right)\mathbf{P}_k^f \\ \mathbf{K}_k &= \mathbf{P}_k^f\,\mathbf{H}_k{}^T\left(\mathbf{H}_k\,\mathbf{P}_k^f\,\mathbf{H}_k{}^T + \mathbf{R}_k\right)^{-1} \end{aligned}$$

equivalent to variationnal approach for KF : minimise cost function $\mathcal{J}$

$$\mathcal{J}(\mathbf{x}) = \frac{1}{2}\left(\mathbf{x} - \mathbf{x}_k^f\right)^T \mathbf{P}_k^{f\,-1}\left(\mathbf{x} - \mathbf{x}_k^f\right) + \frac{1}{2}\left(\mathbf{y}_k^o - \mathcal{H}_k(\mathbf{x})\right)^T \mathbf{R}_k{}^{-1}\left(\mathbf{y}_k^o - \mathcal{H}_k(\mathbf{x})\right)$$

back

Where are we?

The two steps of EnKF

1. **Forecast step**:
$\mathbf{x}^{f}{}^{(i)}_{k+1} = \mathcal{M}_k \left(\mathbf{x}^{a}{}^{(i)}_{k}\right) \ \left(+\mathbf{q}^{(i)}_k\right), i = 1, \ldots, N_{ens}$
with $\mathbf{q}^{(i)}_k$ sampled from model error statistics (not necessary gaussian)
We assume a perfect model: $\mathbf{q}^{(i)}_k = 0$

2. **Analysis step**: various EnKF formulations:
   - Stochastic EnKF [Burgers et al. 1998]: for each member of the ensemble, noisy observation vector (according to observation error statistics)
   - Deterministic EnKF [Bishop et al. 2000, Whitaker Hamill 2001, ...]: observations error statistics used through the covariance matrix

Ensemble Transform Kalman Filter (ETKF)

Detailled in [Hunt et al. 2007], [Harlim and Hunt 2007].

Analysis step

Build an ensemble $\left\{\mathbf{x}^{a(i)}, i = 1 \dots N_{ens}\right\}$ such that $\overline{\mathbf{x}}^a$ is the minimizer of

$$\mathcal{J}(\mathbf{x}) = \frac{1}{2}\left(\mathbf{x} - \overline{\mathbf{x}}^f\right)^T {\mathbf{P}_e^f}^{-1}\left(\mathbf{x} - \overline{\mathbf{x}}^f\right) + \frac{1}{2}\left(\mathbf{y}^o - \mathcal{H}(\mathbf{x})\right)^T \mathbf{R}^{-1}\left(\mathbf{y}^o - \mathcal{H}(\mathbf{x})\right)$$

and $\mathbf{P}_e^a \approx$ inverse of Hessian of $\mathcal{J}$ at the minimum.

Ensemble Transform Kalman Filter (ETKF)

Analysis step

Build an ensemble $\left\{\mathbf{x}^{a(i)}, i = 1 \dots N_{ens}\right\}$ such as $\overline{\mathbf{x}}^a$ minimum of

$$\mathcal{J}(\mathbf{x}) = \frac{1}{2}\left(\mathbf{x} - \overline{\mathbf{x}}^f\right)^T \mathbf{P}_e^{f\,-1}\left(\mathbf{x} - \overline{\mathbf{x}}^f\right) + \frac{1}{2}\left(\mathbf{y}^o - \mathcal{H}(\mathbf{x})\right)^T \mathbf{R}^{-1}\left(\mathbf{y}^o - \mathcal{H}(\mathbf{x})\right)$$

$$\text{and} \qquad \mathbf{P}_e^a \approx \mathbf{P}^a.$$

Attention: $\mathrm{rank}(P_e^f) \leq N_{ens} - 1$ by construction so P_e^f not invertible.

Hypotheses

- Transform Filter

$$\text{Assume } \mathbf{x} = \overline{\mathbf{x}}^f + \mathbf{X}^f\mathbf{w} \text{ and minimise } \mathcal{J}(\overline{\mathbf{x}}^f + \mathbf{X}^f\mathbf{w})$$

  - $\mathbf{w} \in \mathbb{R}^{N_{ens}}$
  - $\mathbf{X}^f$ matrix whose ith column is $\mathbf{x}^{f\,(i)} - \overline{\mathbf{x}}^f$

- Add a term to solve non unique minimum problem
- Linear approximation for observation operator

$$\mathcal{H}\left(\overline{\mathbf{x}}^f + \mathbf{X}^f\mathbf{w}\right) \approx \overline{\mathbf{y}}^f + \mathbf{Y}^f\mathbf{w}$$

with

$$\mathbf{y}^{f\,(i)} = \mathcal{H}\left(\mathbf{x}^{f\,(i)}\right) \qquad \overline{\mathbf{y}}^f = \frac{1}{N}\sum_{i=1}^{N_{ens}} \mathbf{y}^{f\,(i)}$$

matrix $\mathbf{Y}^f$ whose ith column is $\mathbf{y}^{f\,(i)} - \overline{\mathbf{y}}^f$

Minimisation of new cost function

Finally, we search to minimise a quadratic cost function

$$\widetilde{\mathcal{J}}^*(\mathbf{w}) = \frac{N-1}{2}\mathbf{w}^T\mathbf{w} + \frac{1}{2}\left(\mathbf{y}^o - \overline{\mathbf{y}}^f - \mathbf{Y}^f\mathbf{w}\right)^T \mathbf{R}^{-1}\left(\mathbf{y}^o - \overline{\mathbf{y}}^f - \mathbf{Y}^f\mathbf{w}\right)$$

We have a direct computation of:

- minimum of $\widetilde{\mathcal{J}}^*$

$$\overline{\mathbf{w}}^a = \widetilde{\mathbf{P}}^a\,{\mathbf{Y}^f}^T\mathbf{R}^{-1}\left(\mathbf{y}^o - \overline{\mathbf{y}^f}\right)$$

- associated error covariance matrix $\left(= Hess\left(\widetilde{\mathcal{J}}^*(\overline{\mathbf{w}}^a)\right)^{-1}\right)$

$$\widetilde{\mathbf{P}}^a = \left((N_{ens}-1)\,\mathbf{I}_{N_{ens}} + {\mathbf{Y}^f}^T\mathbf{R}^{-1}\mathbf{Y}^f\right)^{-1}$$

- symetric square root matrix

$$\mathbf{W}^a = \left((N_{ens}-1)\widetilde{\mathbf{P}}^a\right)^{1/2}$$

ETKF analysis step

Analysis ensemble $\left\{\mathbf{x}^{a(i)}, i = 1 \dots N_{ens}\right\}$ is defined as

- $\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^f + \mathbf{X}^f \overline{\mathbf{w}}^a$ (mean)
- $\mathbf{X}^a = \mathbf{X}^f \mathbf{W}^a$ (anomalies matrix)
- $\mathbf{x}^{a(i)} = \overline{\mathbf{x}}^a + \mathbf{X}^a_i$ (with $\mathbf{X}^a_i$ ith column of $\mathbf{X}^a$)
- $\mathbf{P}^a_e = \mathbf{X}^f \widetilde{\mathbf{P}}^a {\mathbf{X}^f}^T$

back