Didier Auroux

Université Côte d'Azur, Nice, France

didier.auroux@univ-cotedazur.fr

Observers for data assimilation and parameter estimation

Control Theory and Inverse Problems

Monastir, May 8-10, 2023

Talk overview

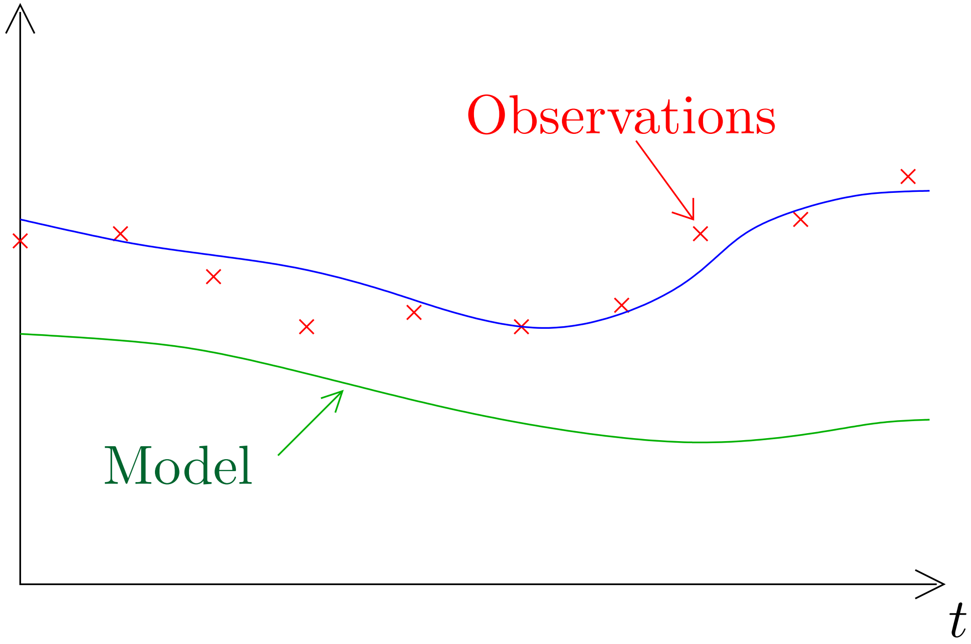
1. Nudging and observers

2. Back and Forth Nudging algorithm

3. Diffusive BFN algorithm

4. Parameter estimation

Data assimilation



combination

model + observations

$\Downarrow$

identification of the state (and/or parameters) of a geophysical system

- 4D-VAR : optimal control method, based on the minimization of the discrepancy between the model solution and the observations.
- Sequential methods : Kalman filtering, ensemble Kalman filters, ...
- Hybrid methods : En-4DVar, 4D-EnVar, ...
- Observer approach : Nudging, Back and Forth Nudging, more complex observers, ...

⇒ 1. Nudging and observers

2. Back and Forth Nudging algorithm

3. Diffusive BFN algorithm

4. Parameter estimation

Forward nudging

Let us consider a model governed by a system of ODE :

$$\frac{dX}{dt} = F(X), \quad 0 < t < T,$$

with an initial condition $X(0) = x_0$.

$\mathcal{Y}(t)$: observations of the system

H : observation operator.

$$\begin{cases} \dfrac{dX}{dt} = F(X) + K(\mathcal{Y} - H(X)), \quad 0 < t < T, \\ X(0) = X_0, \end{cases}$$

where K is the nudging (or gain) matrix.

In the linear case (where F is a matrix), the forward nudging is called Luenberger or asymptotic observer.

Forward nudging

— Meteorology : Hoke-Anthes (1976)

— Oceanography (QG model) : De Mey et al. (1987), Verron-Holland (1989)

— Atmosphere (meso-scale) : Stauffer-Seaman (1990)

— Optimal determination of the nudging coefficients :

Zou-Navon-Le Dimet (1992), Stauffer-Bao (1993),
Vidard-Le Dimet-Piacentini (2003)
Lakshmivarahan-Lewis (2011)

Variational interpretation

Model equation with nudging :

$$\frac{dX}{dt} = FX + K(\mathcal{Y} - HX)$$

Implicit discretization, with X^n at time n and X^{n+1} at time $n+1$:

$$\frac{X^{n+1} - X^n}{\Delta t} = FX^{n+1} + K(\mathcal{Y}^{n+1} - HX^{n+1}).$$

X^{n+1} is solution of the following equation :

$$X - X^n = \Delta t\, FX + \Delta t\, K(\mathcal{Y}^{n+1} - HX)$$

Assume that

$$K = k\, H^T R^{-1}$$

where R is the covariance matrix of the errors of observations.

Variational interpretation

$$X - X^n = \Delta t\, F X + k\, \Delta t\, H^T R^{-1} (\mathcal{Y}^{n+1} - H X)$$

Assume the model F is derived from an energy principle.

Variational interpretation : direct nudging is a compromise between the minimization of the energy of the system and the quadratic distance to the observations :

$$\min_X \left[\frac{1}{2} \langle X - X^n, X - X^n \rangle - \frac{\Delta t}{2} \langle F X, X \rangle + k\, \frac{\Delta t}{2} \langle R^{-1} (\mathcal{Y}^{n+1} - H X), \mathcal{Y}^{n+1} - H X \rangle \right],$$

Example :

Heat equation $F = \Delta$ (Laplacian), and the energy is $-\langle \Delta X, X \rangle = \|\nabla X\|^2$.

Sequential interpretation

It is also possible to give a sequential interpretation of the standard nudging algorithm by seeing it as a Kalman filter. Indeed, when no observations are available, the nudging method simply consists of solving the model equations, like Kalman filters.

On the other hand, when some observations are available, in both nudging and Kalman filters, the model solution is corrected by the innovation vector, i.e. the difference between the observations and the corresponding model state.

If at any time, the nudging matrices are set in an optimal way, then the standard nudging method is equivalent to the standard Kalman filter. In the other cases, it can be seen as a suboptimal Kalman filter.

Nudging : convergence in the linear case

Luenberger observer, or asymptotic observer [Luenberger, 1966]

$$\begin{cases} \dfrac{dX_{true}}{dt} = FX_{true}, \quad \mathcal{Y} = HX_{true}, \\ \dfrac{dX}{dt} = FX + K(\mathcal{Y} - HX). \end{cases}$$

$$\frac{d}{dt}(X - X_{true}) = (F - KH)(X - X_{true})$$

If $F - KH$ is a Hurwitz matrix, i.e. its spectrum is strictly included in the half-plane $\{\lambda \in \mathbb{C}; Re(\lambda) < 0\}$, then $X \to X_{true}$ when $t \to +\infty$.

Pole assignment method : (or pole placement) [Arnold and Datta 1988]
If a system (F, H) is observable, then there exists a matrix K such that $F - KH$ is stable, i.e. all eigenvalues have a strictly negative real part.

$\Rightarrow$ in this case, it is possible to find a nudging matrix K that makes the nudging solution converge towards the true state.

Nudging : linear case example

Simple example : consider the following ODE in $\mathbb{R}^2$:

$$\dot{x}(t) = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} x(t)$$

The characteristic polynomial of the model matrix F is :

$$det(\lambda I - F) = (\lambda - 1)^2 - 1 = \lambda^2 - 2\lambda$$

and the eigenvalues are 0 and 2.

We assume that the first component of x is observed : $H = (1 \quad 0)$. Then we are looking for a gain matrix

$$K = \begin{pmatrix} k_1 \\ k_2 \end{pmatrix}$$

such that $F - KH$ has negative eigenvalues.

Nudging : linear case example

$$F - KH = \begin{pmatrix} 1-k_1 & 1 \\ 1-k_2 & 1 \end{pmatrix}$$

and the characteristic polynomial is :

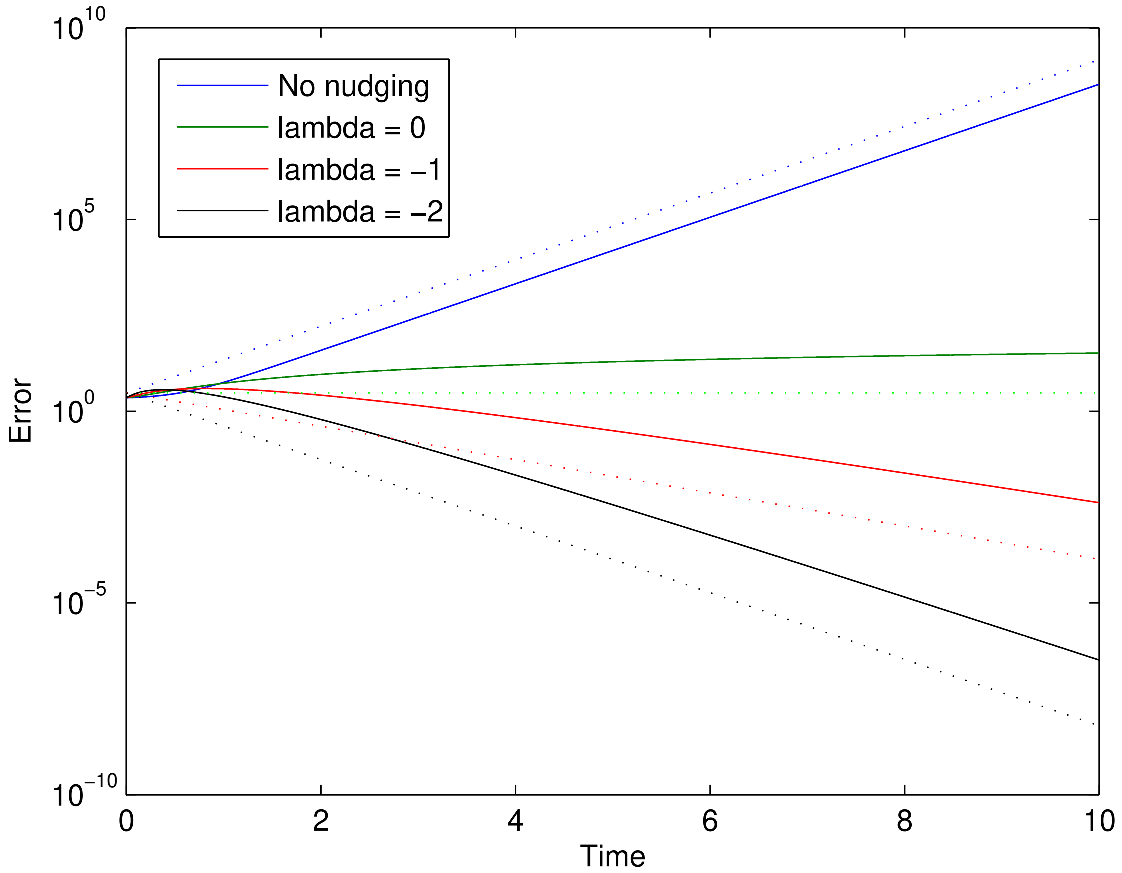
$$det(\lambda I - (F - KH)) = (\lambda - 1 + k_1)(\lambda - 1) - 1 + k_2 = \lambda^2 + (k_1 - 2)\lambda + (k_2 - k_1).$$

Choose for instance $k_1 = 4$ and $k_2 = 5$, which leads to the polynomial $\lambda^2 + 2\lambda + 1$, and the eigenvalues of $F - KH$ are now -1 and -1 : they are both strictly negative (or of strictly negative real part), and the nudging system is now stable.

In this case, the error asymptotically decreases in time : $E(t) = e^{-t} E_0$.

Nudging : linear case example

Numerical tests on this example : $X_{true,0} = [1; -2]$, $X_0 = [1; 0]$.



Case 1 : $k_1 = 2$ and $k_2 = 2 \Rightarrow$ eigenvalues 0 and 0 ;

Case 2 : $k_1 = 4$ and $k_2 = 5 \Rightarrow$ eigenvalues -1 and -1 ;

Case 3 : $k_1 = 6$ and $k_2 = 10 \Rightarrow$ eigenvalues -2 and -2.

Nudging : non-linear case

Luenberger observer for nonlinear models :

We consider a reference trajectory of a nonlinear system given by

$$\frac{dX_{true}}{dt} = FX_{true} + G(X_{true}),$$

where F is the linear part of the model, and G is a nonlinear function, assumed to be differentiable and Lipschitz :

$$\|G(X_1) - G(X_2)\| \leq L\|X_1 - X_2\|, \ \ \forall\ X_1, X_2,$$

where $L > 0$ is a Lipschitz constant.

We assume that the system is observed : $\mathcal{Y} = HX_{true}$, and that (F, H) is observable. Then we know that there exists a matrix K such that $F - KH$ is a stable matrix.

Error equation

We introduce the following observer equation :

$$\frac{dX}{dt} = FX + G(X) + K(\mathcal{Y} - HX) = FX + G(X) + KH(X_{true} - X).$$

This is the standard Luenberger observer (or nudging method).

Let $E = X - X_{true}$ be the error (difference between the observer and true trajectories). Then E satisfies the following equation :

$$\frac{dE}{dt} = FX + G(X) + KH(X_{true} - X) - FX_{true} - G(X_{true})$$

$$= (F - KH)E + G(X) - G(X_{true})$$

Error decrease

$$\frac{1}{2}\frac{d\|E\|^2}{dt} = E.\frac{dE}{dt} = E.\left((F-KH)E\right) + E.\left(G(X)-G(X_{true})\right)$$
$$\leq \lambda_{max}\|E\|^2 + \|E\|\,\|G(X)-G(X_{true})\| \leq (\lambda_{max}+L)\|E\|^2.$$

Then if K is chosen such that all the eigenvalues of $F-KH$ are (of real part) strictly smaller than $-L$, the opposite of the Lipschitz constant of G, then the square norm of the error decreases asymptotically in time.

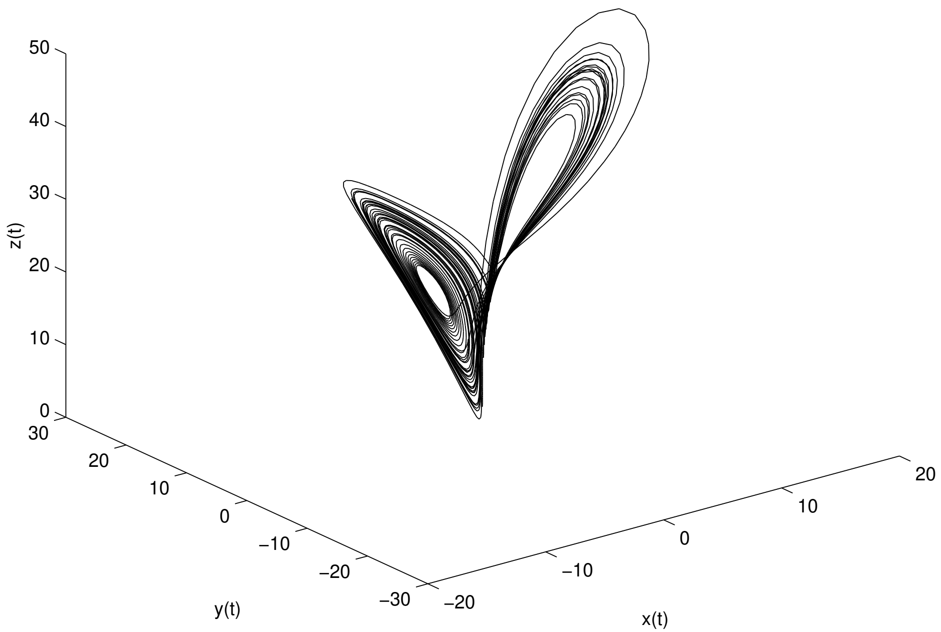
And then $X \to X_{true}$ when time goes to infinity.

Non-linear example

Example : Lorenz system

$$\begin{cases} \dot{x} = \sigma(y - x), \\ \dot{y} = \rho x - y - xz, \\ \dot{z} = xy - \beta z, \end{cases}$$

with standard values of parameters $\sigma = 10$, $\rho = 28$ and $\beta = \frac{8}{3}$ for a chaotic behavior.



Lorenz : model decomposition

The model can be decomposed in a linear part, and a nonlinear part :

$$\frac{dX}{dt} = FX + G(X),$$

with

$$X = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \qquad F = \begin{pmatrix} -\sigma & \sigma & 0 \\ \rho & -1 & 0 \\ 0 & 0 & -\beta \end{pmatrix}, \qquad G(X) = \begin{pmatrix} 0 \\ -xz \\ xy \end{pmatrix}.$$

Assuming that all considered trajectories are bounded (see previous figure), then the function G is Lipschitz.
Then we can define the following Luenberger observer :

$$\frac{dX}{dt} = FX + G(X) + KH(X_{true} - X).$$

Lorenz : observability conditions

If we choose $H = (1\ ;\ 0\ ;\ 1)$:
Then $HF = (-\sigma\ ;\ \sigma\ ;\ -\beta)$, and $HF^2 = (\sigma^2 + \rho\sigma\ ;\ -\sigma^2 - \sigma\ ;\ \beta^2)$.

As the matrix $(H; HF; HF^2)$ is invertible, then (F, H) is observable, and we can place the poles of F : there exists K such that $F - KH$ is stable.

With $H = (1\ ;\ 0\ ;\ 0)$ or $H = (0\ ;\ 1\ ;\ 0)$, the system is not observable, and then it is not possible to place the poles anywhere, but it is still possible to find K such that $F - KH$ is stable.

For instance with $H = (1\ ;\ 0\ ;\ 0)$ and $K = (0\ ;\ \rho\ ;\ 0)^T$,

$$F - KH = \begin{pmatrix} -\sigma & \sigma & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -\beta \end{pmatrix}$$

3 negative eigenvalues : -1, $-\beta$ and $-\sigma$.

Lorenz : non-linear observer

Another possible observer :

$$\begin{cases} \dot{x} = \sigma(y - x), \\ \dot{y} = \rho x_{true} - y - x_{true} z, \\ \dot{z} = x_{true} y - \beta z, \end{cases}$$

which can be rewritten as

$$\begin{cases} \dot{x} = \sigma(y - x), \\ \dot{y} = \rho x - y - xz + \rho(x_{true} - x) - z(x_{true} - x), \\ \dot{z} = xy - \beta z + y(x_{true} - x), \end{cases}$$

Only x is observed, and compared to the previous example with $K = (0; \rho; 0)^T$, there is an additional term : $(0\ ;\ -z\ ;\ y)^T H(X_{true} - X)$

Lorenz : non-linear observer

Then the error $E = X - X_{true}$ is solution of the following ODE :

$$\frac{dE}{dt} = (F - KH)E + S(t)E,$$

with

$$F - KH = \begin{pmatrix} -\sigma & \sigma & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -\beta \end{pmatrix},$$

and where $S(t)$ is the following matrix :

$$S(t) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -x_{true}(t) \\ 0 & x_{true}(t) & 0 \end{pmatrix}$$

Spectral decomposition

Let us compute the eigenvalues of $F - KH - S(t)$: the characteristic polynomial is

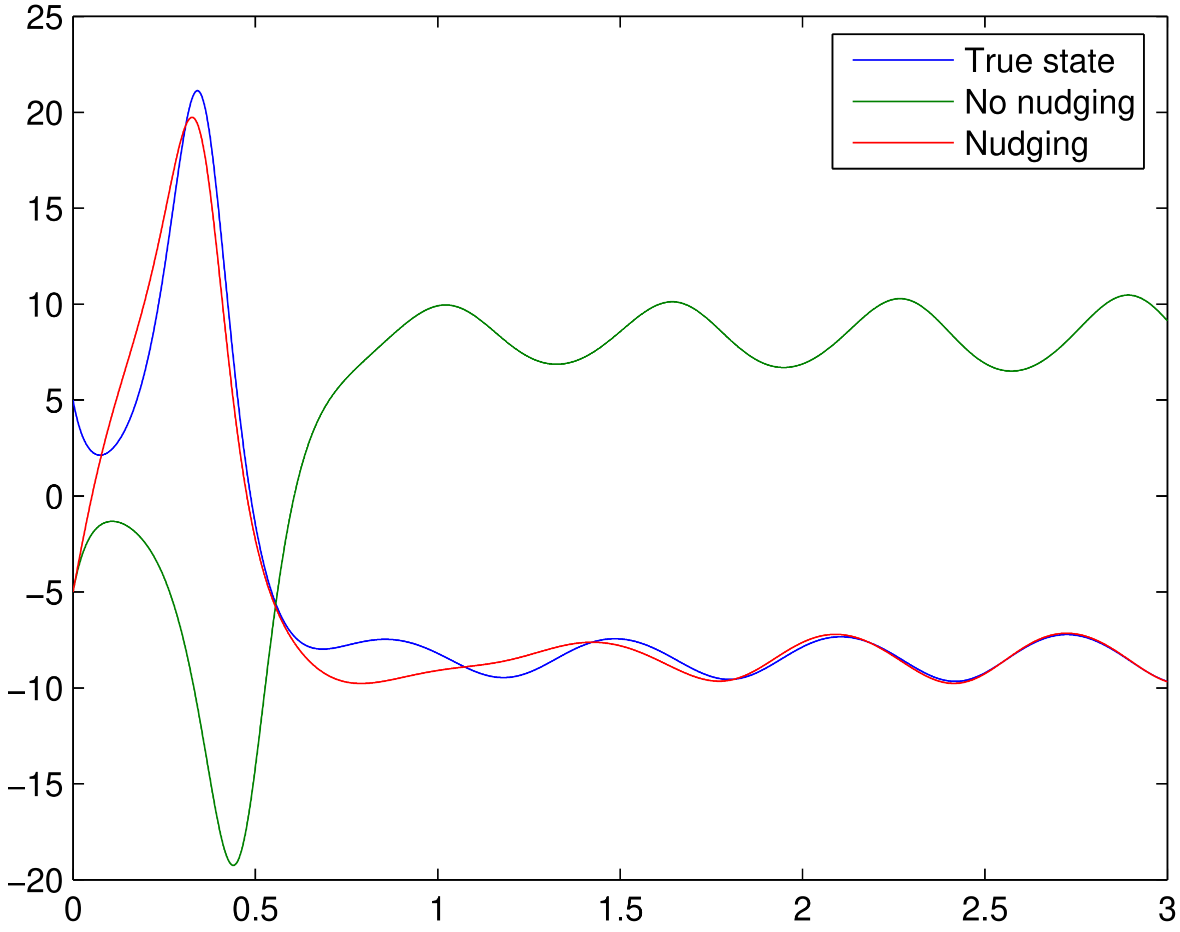
$$(\lambda + \sigma)(\lambda^2 + \lambda(1 + \beta) + (\beta + x_{true}(t)^2).$$

As $\beta > 0$ and $\sigma > 0$, for any value of $x_{true}(t)$, all three eigenvalues are strictly negative, or their real parts are strictly negative, and then the error decreases asymptotically in time.

Numerical tests

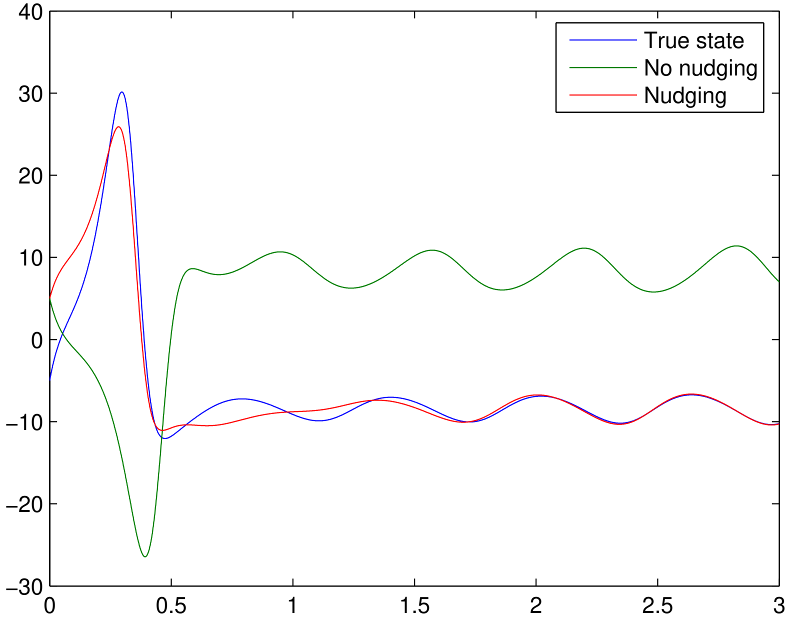
Numerical experiment on this example :

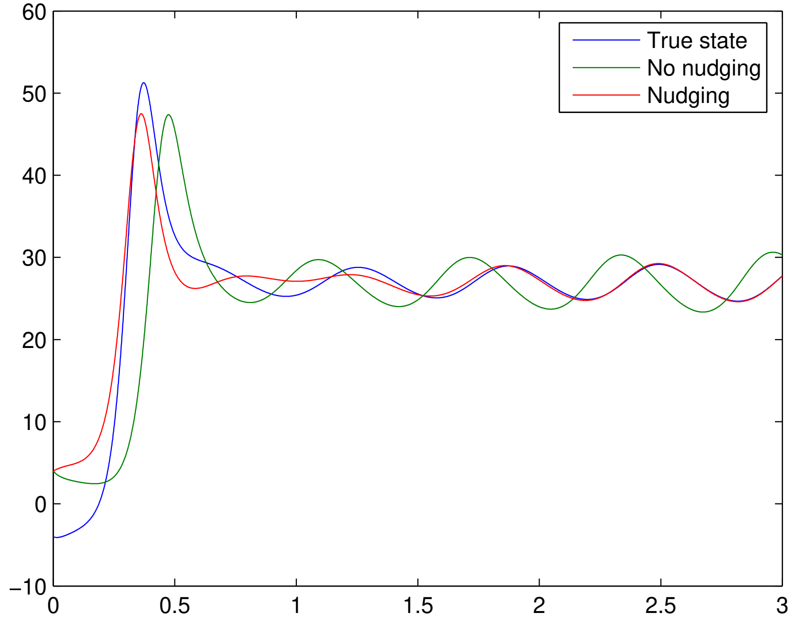
$X_{true}(0) = [5; -5; -4]$ and $X(0) = [-5; 5; 4]$.



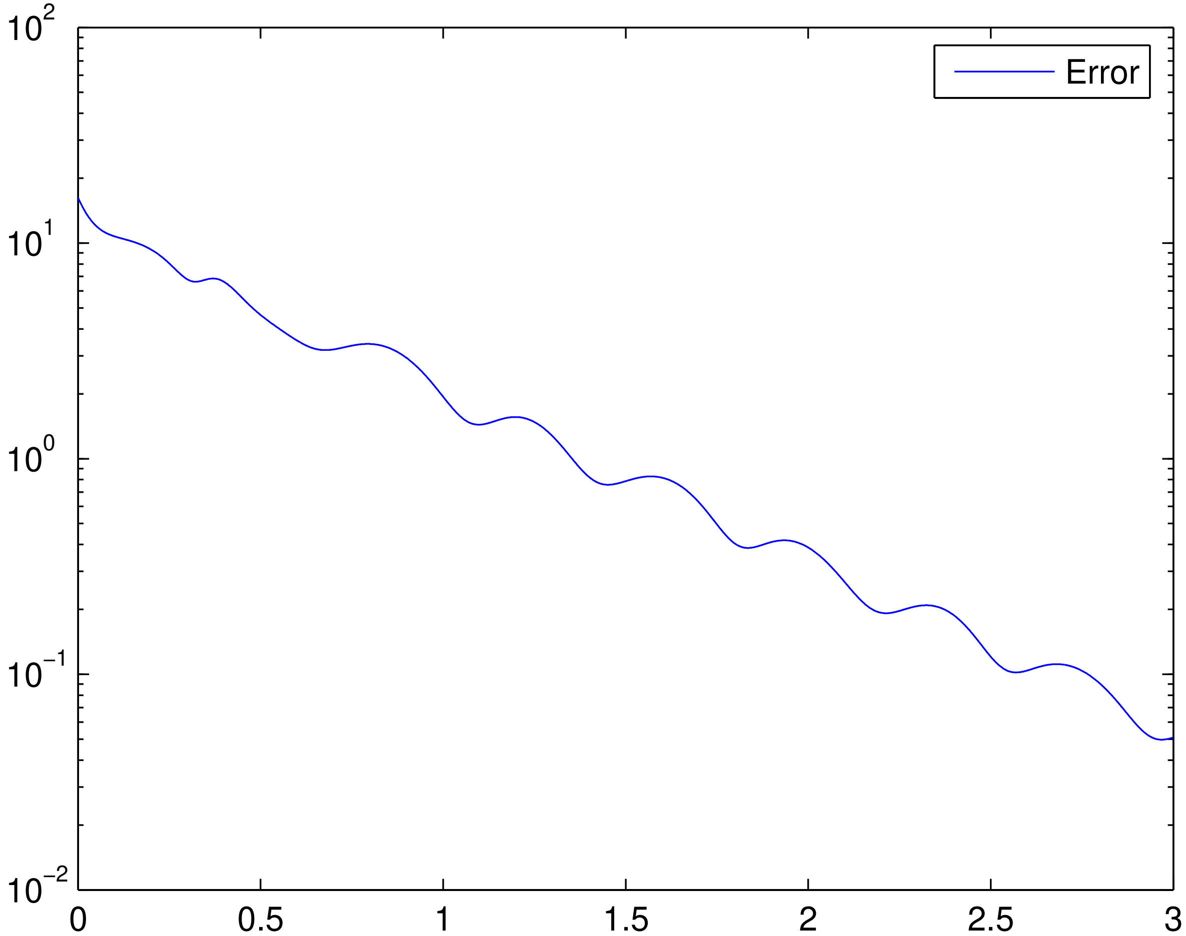
Numerical tests

Convergence of y and z also :





Numerical tests



Evolution of the norm of the error between $X_{nudging}$ and X_{true} versus time (largest eigenvalue $\simeq -1$).

More complex observers

If the model has more than one variable (or if all components are not observed), the standard nudging only corrects the observed variables with themselves.

$\Rightarrow$ extension to more complex observers, in which non observed variables are controlled by observed ones.

Example on a 2D shallow water model :

$$\begin{cases} \dfrac{\partial h}{\partial t} = -\nabla \cdot (hv), \\ \dfrac{\partial v}{\partial t} = -(v \cdot \nabla)v - g\nabla h \end{cases}$$

on a square domain with rigid boundaries and no-slip lateral boundary conditions. These equations are derived from Navier-Stokes equations, assuming the horizontal scale is much greater than the vertical one $\Rightarrow$ conservation of mass and of momentum.

Can we identify/correct both variables (height and velocity) if only the water height h is observed ?

Observer design

Any non-linear observer for this model writes :

$$\begin{cases} \dfrac{\partial h}{\partial t} = -\nabla \cdot (hv) + F_h(h_{obs}, v, h), \\ \dfrac{\partial v}{\partial t} = -(v \cdot \nabla)v - g\nabla h + F_v(h_{obs}, v, h), \end{cases}$$

where $F = 0$ when the estimated height h is equal to the observed height h_{obs}.

Formal requirements : symmetry preservation (invariance to translations and rotations of the model, and then of the observer), smoothing by convolution (noisy data), local stability (strong asymptotic convergence of the linearized error system)

Most simple observer that should work : (smallest order of derivative)

$$F_h = \varphi_h * (h - h_{obs}), \qquad F_v = \varphi_v * \nabla(h - h_{obs})$$

with simple invariant kernels :

$$\varphi(x, y) = \beta \exp(-\alpha(x^2 + y^2)).$$

Convergence study

Convergence on the linearized system : let δh and δv be the perturbations around the reference state, and let $\tilde{h} = \delta h - \delta h_{true}$ and $\tilde{v} = \delta v - \delta v_{true}$ be the estimation errors, solutions of

$$\frac{\partial \tilde{h}}{\partial t} = -\bar{h}\, \nabla \cdot \tilde{v} - \varphi_h * \tilde{h},$$

$$\frac{\partial \tilde{v}}{\partial t} = -g \nabla \tilde{h} - \varphi_v * \nabla \tilde{h}.$$

Eliminating $\tilde{v}$ yields a modified damped wave equation with external viscous damping :

$$\frac{\partial^2 \tilde{h}}{\partial t^2} = g\bar{h}\Delta\tilde{h} + \bar{h}\, \varphi_v * \Delta\tilde{h} - \varphi_h * \frac{\partial \tilde{h}}{\partial t}.$$

Convergence study

Theorem : If φ_v and φ_h are defined by $\varphi(x,y) = \beta \exp(-\alpha(x^2+y^2))$ with $\beta_v, \beta_h, \alpha_v, \alpha_h > 0$, then the first order approximation of the error system around the equilibrium $(h,v) = (\bar{h}, 0)$ is strongly asymptotically convergent. Indeed if we consider the following Hilbert space and norm : $\mathcal{H} = H^1(\Omega) \times L^2(\Omega)$,

$$\|(u,w)\|_{\mathcal{H}} = \left(\int_\Omega \|\nabla u\|^2 + |w|^2 \right)^{1/2},$$

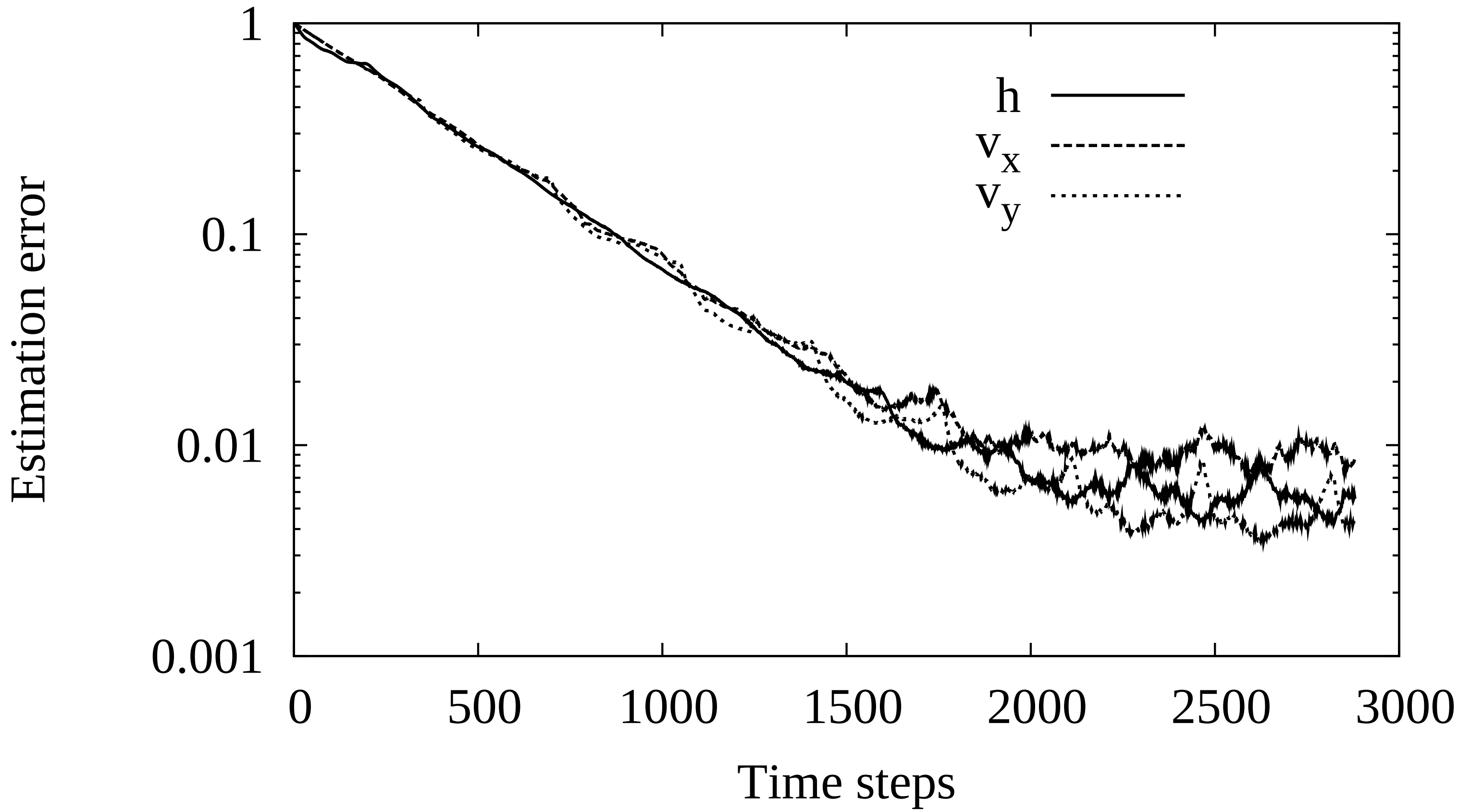
then

$$\lim_{t\to\infty} \left\| \left(\tilde{h}(t), \frac{\partial \tilde{h}}{\partial t}(t) \right) \right\|_{\mathcal{H}} = 0\,.$$

This theorem proves the strong and asymptotic convergence of the error $\tilde{h}$ towards 0, and then it also gives the same convergence for $\tilde{v}$. We deduce that the observer tends to the true state when time goes to infinity.

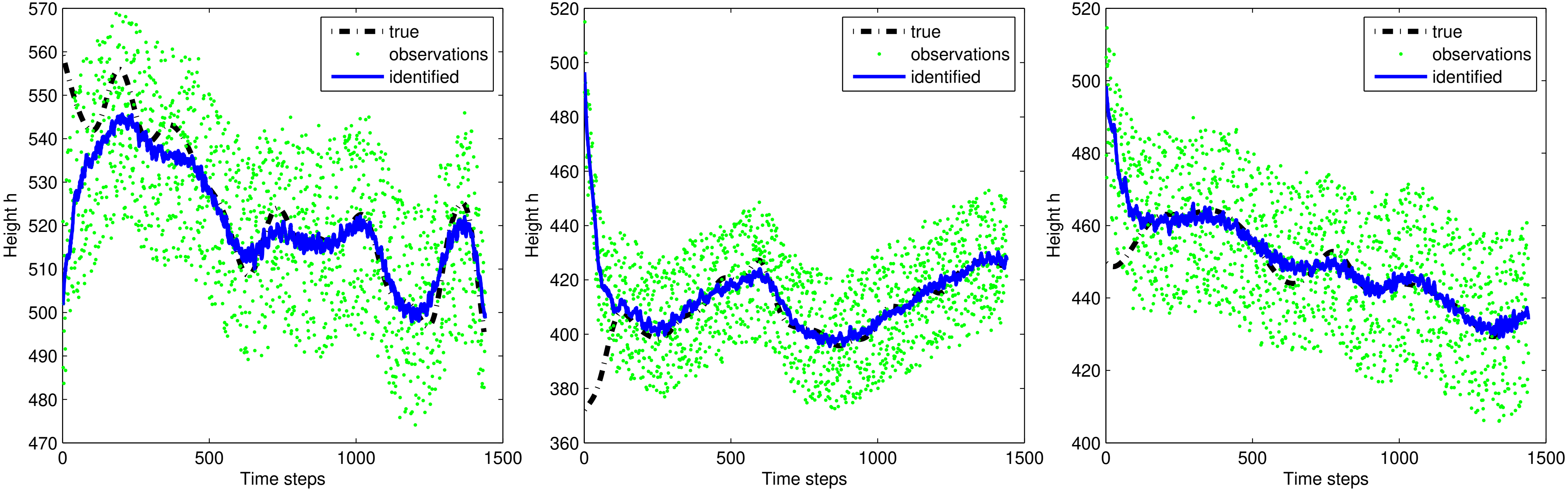
Proof : based on Fourier decomposition of the solution.

Numerical tests



Evolution of the estimation error in relative norm versus the number of time steps, in the case of noisy observations (20% noise), with $\alpha_h = \alpha_v = 1\ m^{-2}$ and $\beta_h = 2.10^{-7}\ s^{-1}$, and with a 100% error on the initial conditions, for the height h, longitudinal velocity v_x and transversal velocity v_y.

Numerical tests : non-linear model



Evolution of the true height, the observed (noisy) height, and the identified (observer) height versus time, for three different points of the domain, located along the energetic current in the middle of the domain.

N-d compressible Navier-Stokes

Compressible Navier-Stokes equation :

$$\begin{cases} \rho_t + \nabla \cdot (\rho u) = 0, \\ \rho[u_t + (u \cdot \nabla) u] = -\nabla p(\rho) + \mu \Delta u + (\lambda + \mu) \nabla (\nabla \cdot u), \end{cases}$$

where $p(\rho) = \rho^\gamma$, $\mu > 0$, $\lambda + \frac{2\mu}{3} \geq 0$.

Space domain : $[0,1]^n$ with periodic conditions ; Time domain : $[0,T]$.

Observers system :

$$\begin{cases} \hat{\rho}_t + \nabla \cdot (\rho \hat{u}) = F_\rho(\hat{\rho}, \hat{u}, u), \\ \rho[\hat{u}_t + (\hat{u} \cdot \nabla) \hat{u}] = -\nabla p(\hat{\rho}) + \mu \Delta \hat{u} + (\lambda + \mu) \nabla (\nabla \cdot \hat{u}) + F_u(\hat{\rho}, \hat{u}, u), \end{cases}$$

with $F_\rho(\hat{\rho}, \hat{u}, u) = \varphi_\rho * D_\rho(u - \hat{u})$, $F_u(\hat{\rho}, \hat{u}, u) = \varphi_u * D_u(u - \hat{u})$, and D_ρ and D_u are differential (or integral) operators.

N-d compressible Navier-Stokes

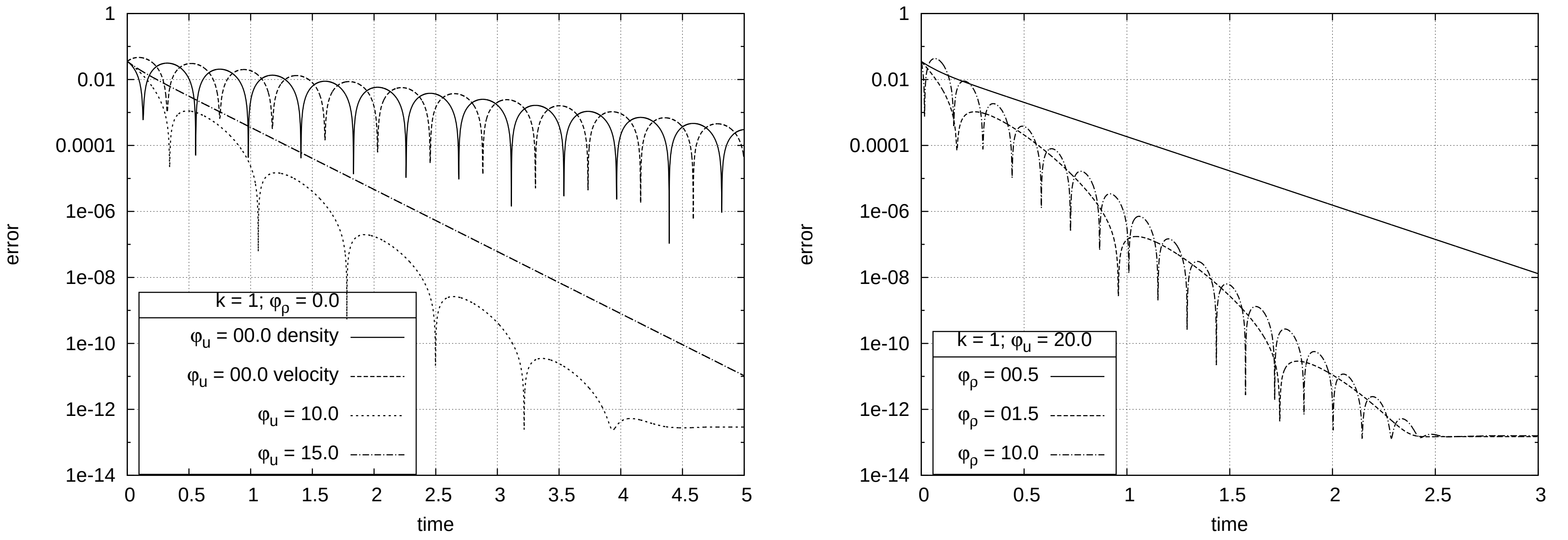
Theorem - linear system : For any $d > 0$, for any $K > 0$, one can find $\varphi_\rho(x)$ and $\varphi_u(x)$ such that the maximal decay rate of the errors on density and velocity towards 0 is at least d for any Fourier mode k such that $|k| \leq K$. The following values can be chosen :

$$\begin{aligned}
\varphi_{u0} &= d\,, \quad \varphi_{\rho 0} = 0\,, && (1)\\
\varphi_{uk} &= \max\{0; d - c_1|k|^2; 2d - (c_1+c_2)|k|^2)\}\,, \quad 0 < |k| \leq K\,, && (2)\\
\varphi_{\rho k} &= \max\left\{0; \frac{\left((c_1+c_2)|k|^2 + \varphi_{uk}\right)^2}{4c_5|k|^2} - 1\right\}\,, \quad 0 < |k| \leq K\,, && (3)\\
\varphi_{uk} &= 0\,, \quad \varphi_{\rho k} = 0\,, \quad |k| > K\,. && (4)
\end{aligned}$$

Proof based on Fourier decomposition of the solutions and spectral analysis of the coupled system. Note that we get explicit Fourier coefficients of the convolution kernels.

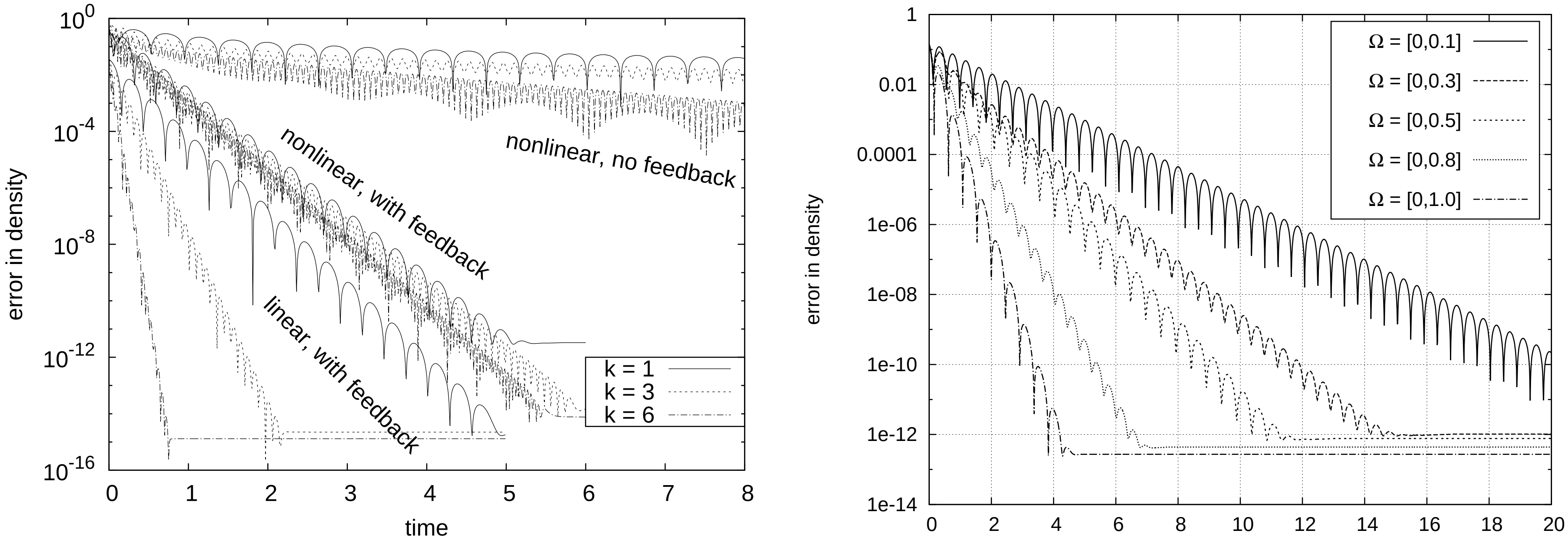
Similar results with density observations, and explicit decay rate calculation in case of nudging (no convolution, no correction of non-observed variables)

Numerical results on 1-d CNS



L^2 norm of the difference between the observer $(\hat{\rho}, \hat{u})$ and the solution (ρ, u) versus time. Solid and dotted lines are the errors in ρ and u, respectively. The left panel is for fixed $\varphi_\rho = 0$ with varying φ_u while the right panel is for fixed $\varphi_u = 20$ with varying φ_ρ.

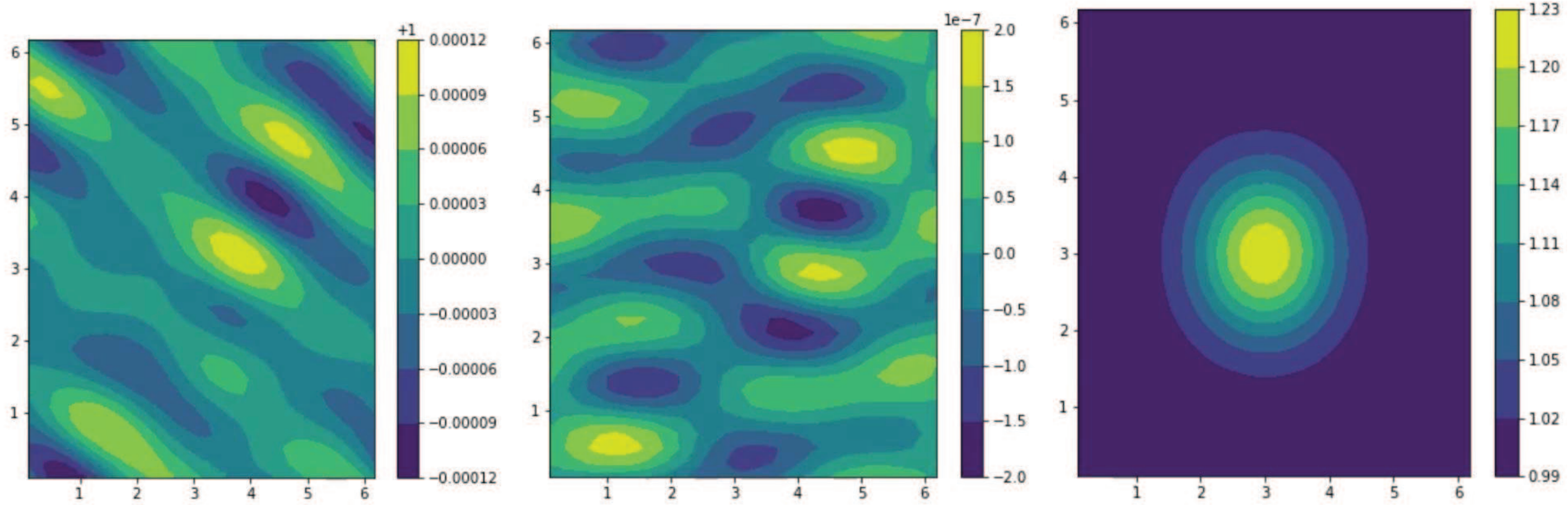
Numerical results on 1-d CNS



The decay rates of the nonlinear observer compared with that of the linear observer (left). The decay rates for the nonlinear observer with observations over partial domain (right).

Image driven by CNS

Observer design for tracking an image driven by compressible Navier-Stokes :



Density, velocity, and passive tracer (image)

$$\hat\rho_t + \nabla\cdot(\hat\rho\hat u) = F_\rho(\bar I, \hat I)$$

$$\hat\rho[\hat u_t + (\hat u\cdot\nabla)\hat u] = -\gamma\hat\rho^{\gamma-1}\nabla\hat\rho + \mu\Delta\hat u + \lambda_\mu\nabla(\nabla\cdot\hat u) + F_u(\bar I, \hat I)$$

$$\hat I_t + \nabla\cdot(\hat u\hat I) = F_I(\bar I, \hat I).$$

where $\bar I$ is the observed image.

Image driven by CNS

Using the following observers :

$$\begin{aligned} F_\rho(\bar{I},\hat{I}) &= \varphi_\rho * (\bar{I}-\hat{I}) \\ F_u(\bar{I},\hat{I}) &= -\rho_0\varphi_u * \nabla(\bar{I}-\hat{I}) \\ F_I(\bar{I},\hat{I}) &= \varphi_I * (\bar{I}-\hat{I}) \end{aligned}$$

we can study the eigenvalues of the linearized equations :

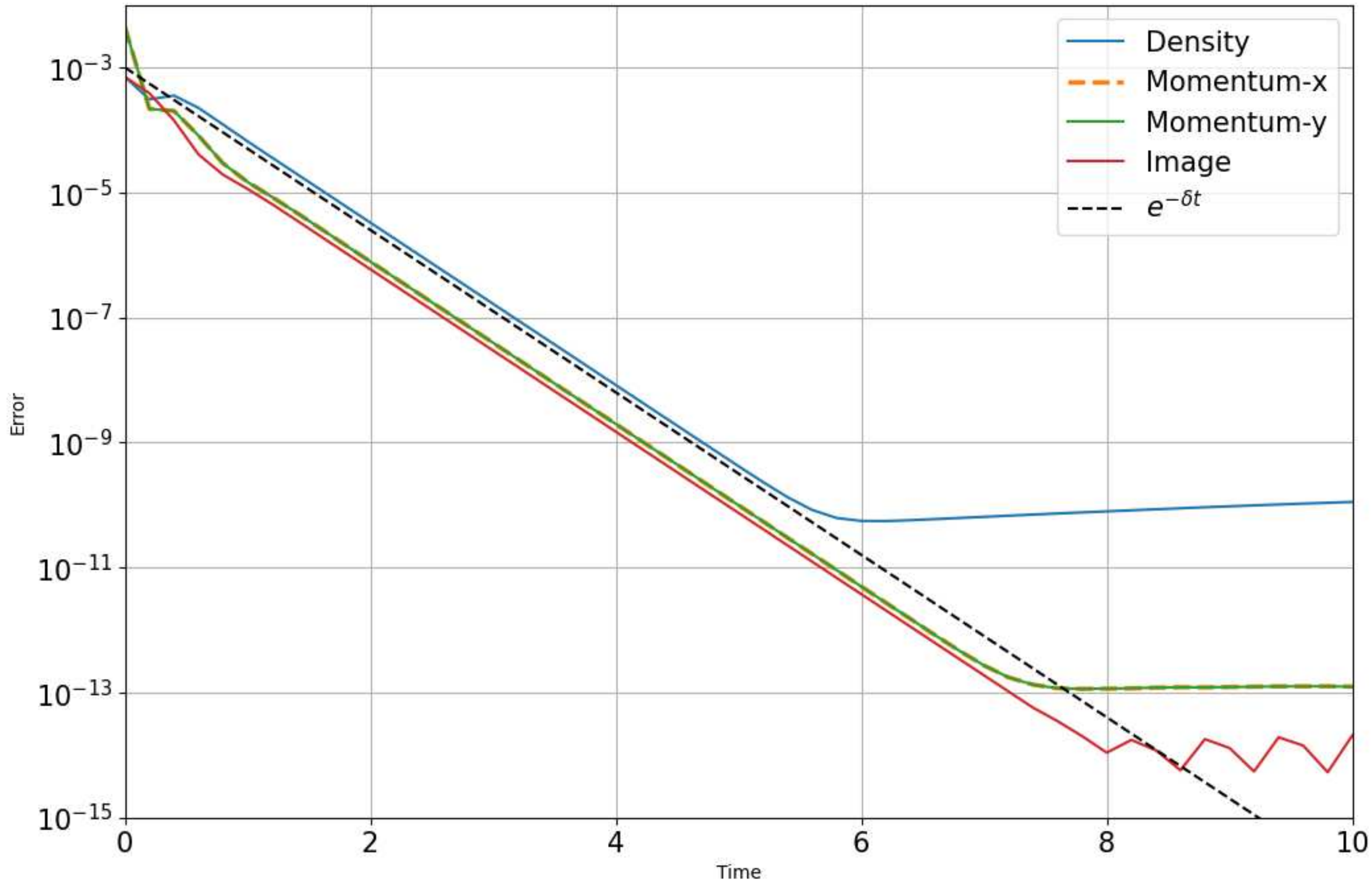
$$\begin{pmatrix} 0 & c_4|k|^2 & \varphi_\rho \\ -c_3 & (-c_1+c_2)|k|^2 & -2\pi i\varphi_u \\ 0 & c_6|k|^2 & -\varphi_I \end{pmatrix}$$

and find optimal Fourier coefficients of the kernels φ in order to control all Fourier modes (up to any maximum mode $|k| \leq K$) at any desired decay rate.

We can also use constant coefficients ($\rightsquigarrow$ standard nudging) and get a minimal decay rate for all Fourier modes.

Image driven by CNS

Example of decay of the error (on all variables) versus time, for a given mode in 2D :



1. Nudging and observers

⇒ 2. Back and Forth Nudging algorithm

3. Diffusive BFN algorithm

4. Parameter estimation

Backward nudging

How can we recover the initial state from the final solution ?

Backward model :

$$\begin{cases} \dfrac{d\tilde{X}}{dt} = F(\tilde{X}), \quad T > t > 0, \\ \tilde{X}(T) = \tilde{X}_T. \end{cases}$$

If we apply nudging to this backward model :

$$\begin{cases} \dfrac{d\tilde{X}}{dt} = F(\tilde{X}) - K(\mathcal{Y} - H\tilde{X}), \quad T > t > 0, \\ \tilde{X}(T) = \tilde{X}_T. \end{cases}$$

BFN : Back and Forth Nudging algorithm

Iterative algorithm (forward and backward resolutions) :

$$\tilde{X}_0(0) = X_b \text{ (first guess)}$$

$$\begin{cases} \dfrac{dX_k}{dt} = F(X_k) + K(\mathcal{Y} - H(X_k)) \\ X_k(0) = \tilde{X}_{k-1}(0) \end{cases}$$

$$\begin{cases} \dfrac{d\tilde{X}_k}{dt} = F(\tilde{X}_k) - K'(\mathcal{Y} - H(\tilde{X}_k)) \\ \tilde{X}_k(T) = X_k(T) \end{cases}$$

If X_k and $\tilde{X}_k$ converge towards the same limit X, and if $K = K'$, then X satisfies the state equation and fits to the observations.

Choice of the direct nudging matrix K

Implicit discretization of the direct model equation with nudging :

$$\frac{X^{n+1} - X^n}{\Delta t} = FX^{n+1} + K(\mathcal{Y} - HX^{n+1}).$$

Variational interpretation : direct nudging is a compromise between the minimization of the energy of the system and the quadratic distance to the observations :

$$\min_X \left[\frac{1}{2} \langle X - X^n, X - X^n \rangle - \frac{\Delta t}{2} \langle FX, X \rangle + \frac{\Delta t}{2} \langle R^{-1}(\mathcal{Y} - HX), \mathcal{Y} - HX \rangle \right],$$

by choosing

$$K = H^T R^{-1}$$

where R is the covariance matrix of the errors of observation.

Choice of the backward nudging K'

The feedback term has a double role :

- stabilization of the backward resolution of the model (irreversible system)
- feedback to the observations

If the system is observable, i.e. $rank[H, HF, \ldots, HF^{N-1}] = N$, then there exists a matrix K' such that $-F - K'H$ is a Hurwitz matrix (pole assignment method).

Simpler solution : one can define $K' = k'H^T R^{-1}$, where k' is e.g. the smallest value making the backward numerical integration stable.

Example of convergence results

Viscous linear transport equation :

$$\begin{cases} \partial_t u - \nu\partial_{xx}u + a(x)\partial_x u = -K(u - u_{obs}), & u(x, t = 0) = u_0(x) \\ \partial_t \tilde{u} - \nu\partial_{xx}\tilde{u} + a(x)\partial_x \tilde{u} = K'(\tilde{u} - u_{obs}), & \tilde{u}(x, t = T) = u_T(x) \end{cases}$$

We set $w(t) = u(t) - u_{obs}(t)$ and $\tilde{w}(t) = \tilde{u}(t) - u_{obs}(t)$ the errors.

- If K and K' are constant, then $\forall t \in [0, T]$: $\widetilde{w}(t) = e^{(-K-K')(T-t)}w(t)$

(still true if the observation period does not cover $[0, T]$)

- If the domain is not fully observed, then the problem is ill-posed.

Error after k iterations : $w_k(0) = e^{-[(K+K')kT]}w_0(0)$

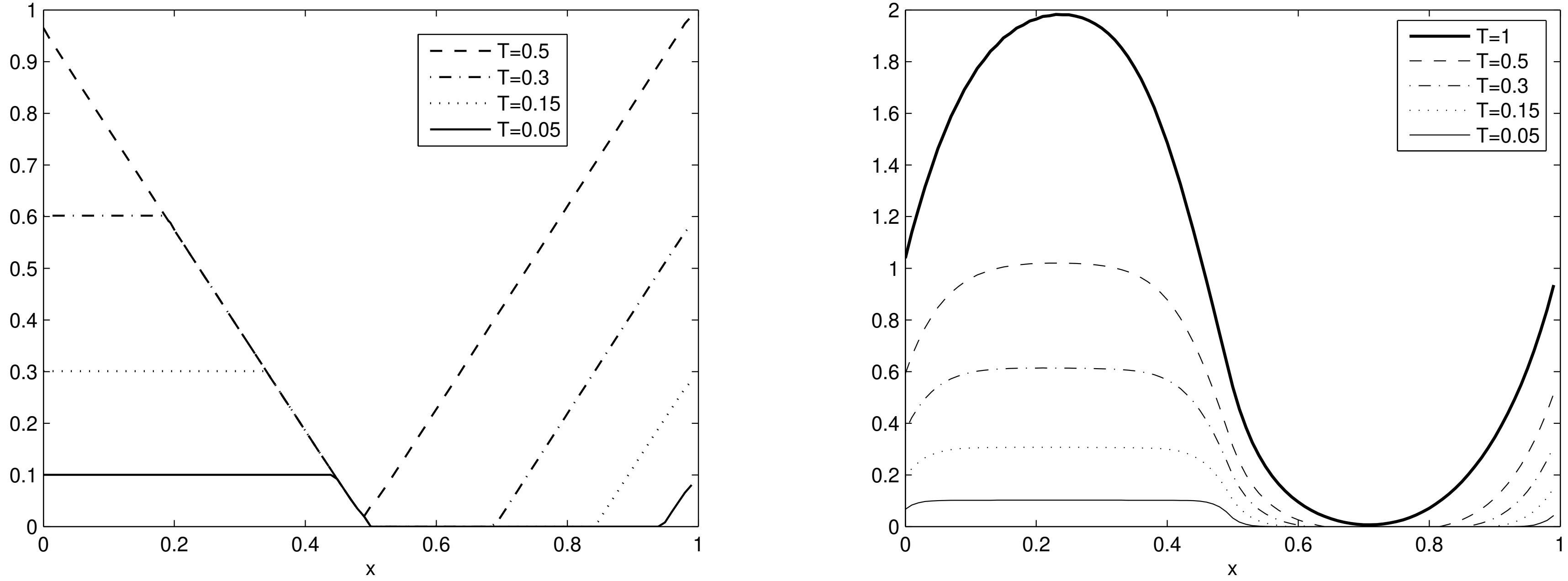
$\rightsquigarrow$ exponential decrease of the error, thanks to :

- $K + K'$: infinite feedback to the observations (not physical)
- T : asymptotic observer (Luenberger)
- k : infinite number of iterations (BFN)

Observability condition

Let $\chi(x)$ be the time during which the characteristic curve wi th foot x lies in the support of K. Then the system is observable if and onl y if $\min_x \chi(x) > 0$.

Partial observations in space : half of the domain is observed.



Decrease rate of the error after one iteration of BFN as a function of the space variable x, for various final times T.

Linear case (left) : theoretical observability condition = $T > 0.5$

Nonlinear case (right) : numerical observability condition = $T > 1$

Shallow water model

$$
\begin{aligned}
\partial_t u - (f+\zeta)v + \partial_x B &= \frac{\tau_x}{\rho_0 h} - ru + \nu\Delta u \\
\partial_t v + (f+\zeta)u + \partial_y B &= \frac{\tau_y}{\rho_0 h} - rv + \nu\Delta v \\
\partial_t h + \partial_x(hu) + \partial_y(hv) &= 0
\end{aligned}
$$

- $\zeta = \partial_x v - \partial_y u$ is the relative vorticity ;
- $B = g^* h + \frac{1}{2}(u^2 + v^2)$ is the Bernoulli potential ;
- $g^* = 0.02\ m.s^{-2}$ is the reduced gravity ;
- $f = f_0 + \beta y$ is the Coriolis parameter (in the β-plane approximation), with $f_0 = 7.10^{-5}\ s^{-1}$ and $\beta = 2.10^{-11}\ m^{-1}.s^{-1}$;
- $\tau = (\tau_x, \tau_y)$ is the forcing term of the model (e.g. the wind stress), with a maximum amplitude of $\tau_0 = 0.05\ s^{-2}$;
- $\rho_0 = 10^3\ kg.m^{-3}$ is the water density ;
- $r = 9.10^{-8}\ s^{-1}$ is the friction coefficient.
- $\nu = 5\ m^2.s^{-1}$ is the viscosity (or dissipation) coefficient.

Shallow water model

2D shallow water model, state = height h and horizontal velocity (u, v)

Numerical parameters :

Domain : $L = 2000$ km $\times$ 2000 km ;
Rigid boundary and no-slip BC ;
Time step = 1800 s ;
Assimilation period : 15 days ;
Forecast period : 15 + 45 days

(run example)

Observations : of h only ($\sim$ satellite obs),
every 5 gridpoints in each space direction,
every 24 hours.

Background : true state one month before the beginning of the assimilation period + white gaussian noise ($\sim$ 10%)

Shallow water model

BFN
(5 iter.)

4DVAR
(5 iter.)

(BFN vs 4DVAR)

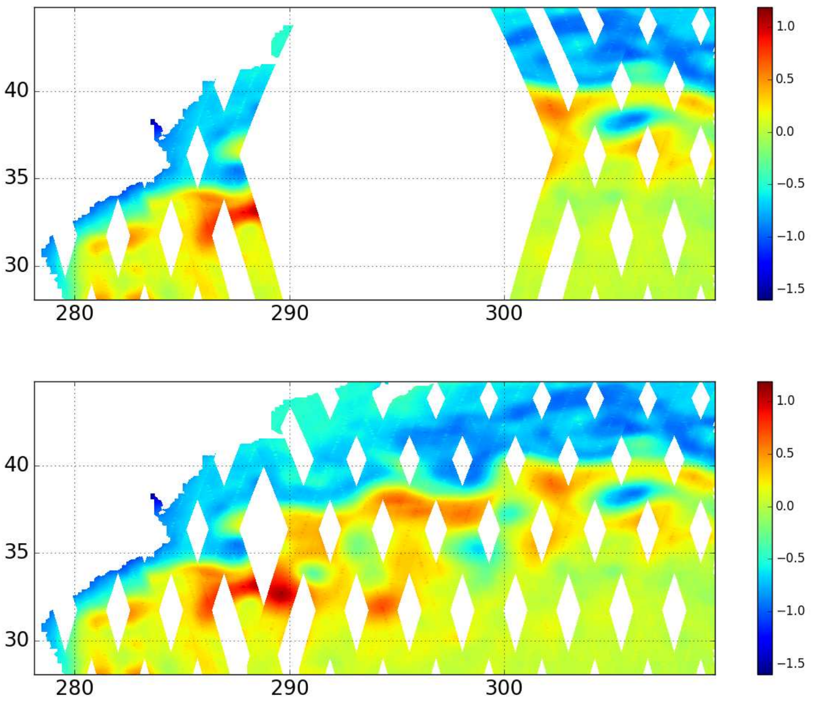
True
state

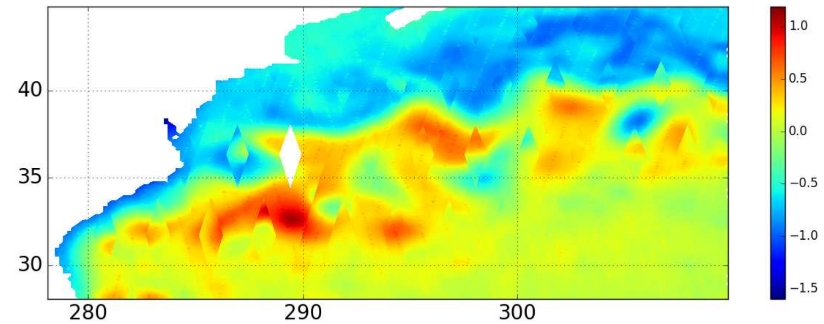
Back-
ground

Comparison BFN - 4DVAR : height h

SWOT observations on a QG model

SWOT (Surface Water Ocean Topography) satellite mission (expected to be launched in 4 months), expected to provide SSH with a swath wide of 120km and repeat period of 21 days

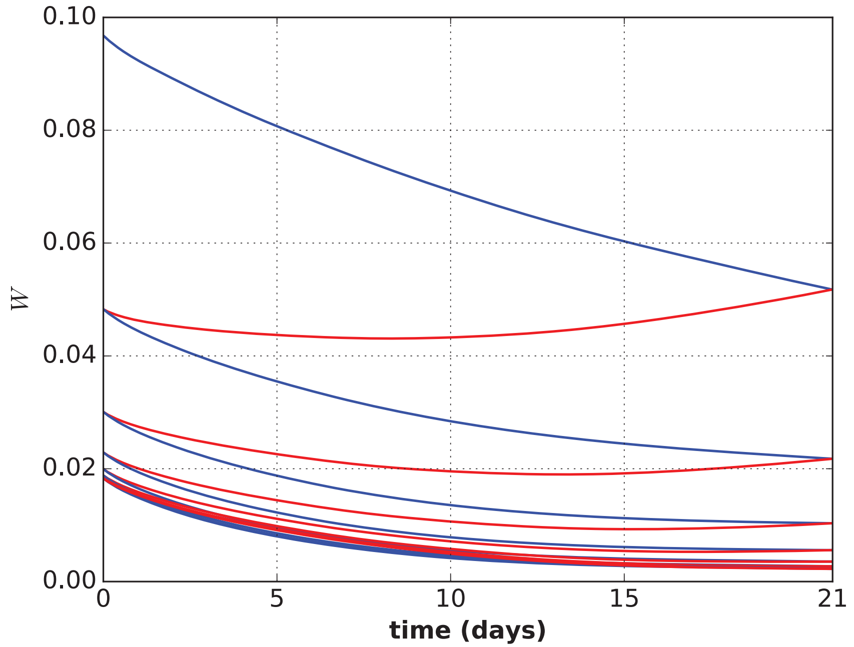


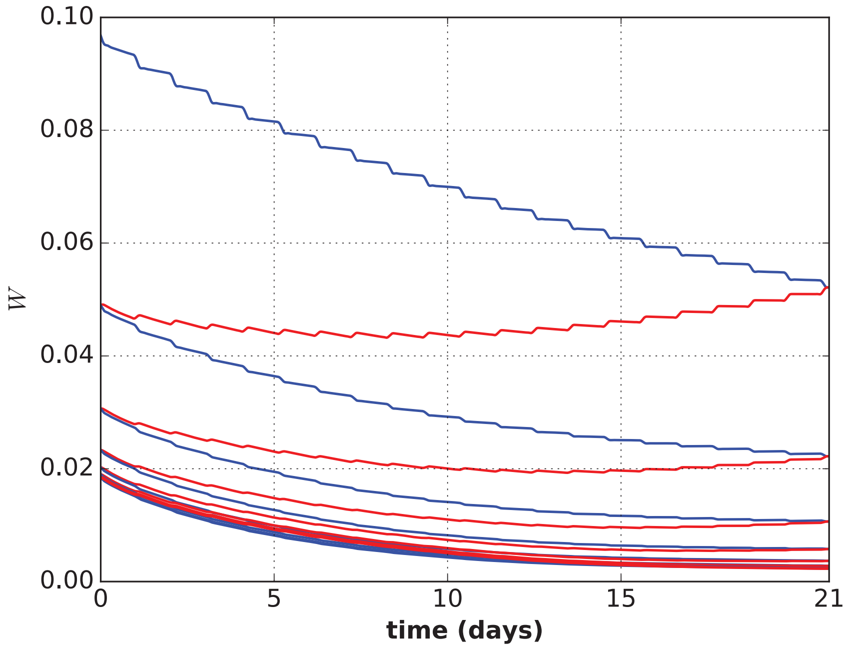


SWOT satellite SSH data coverage after 5, 10 and 21 days.

SWOT observations on a QG model

Thanks to LaSalle's invariance principle, we can define a Lyapunov function that asymptotically decreases towards 0 when time goes to infinity. This ensures the theoretical convergence of the BFN. Moreover, the convergence rate is not impacted by data time frequency :

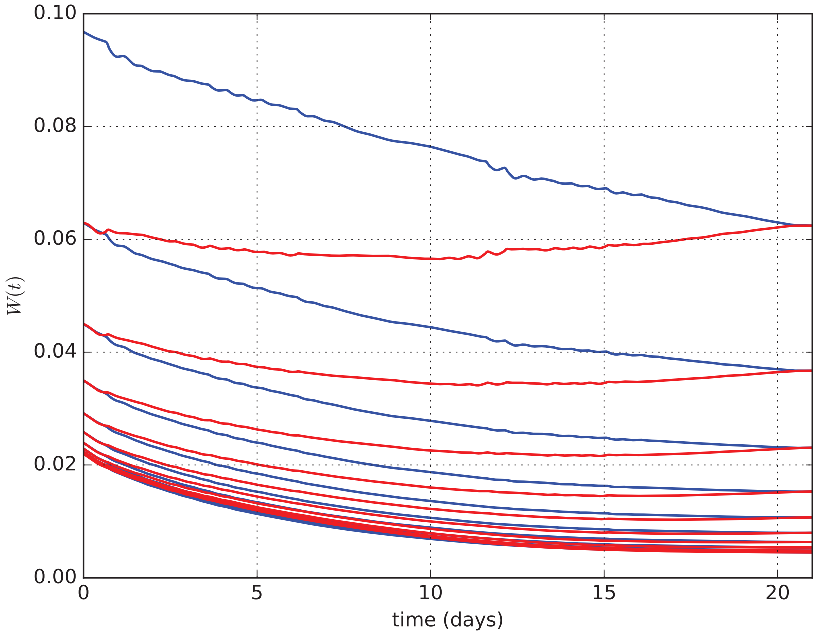


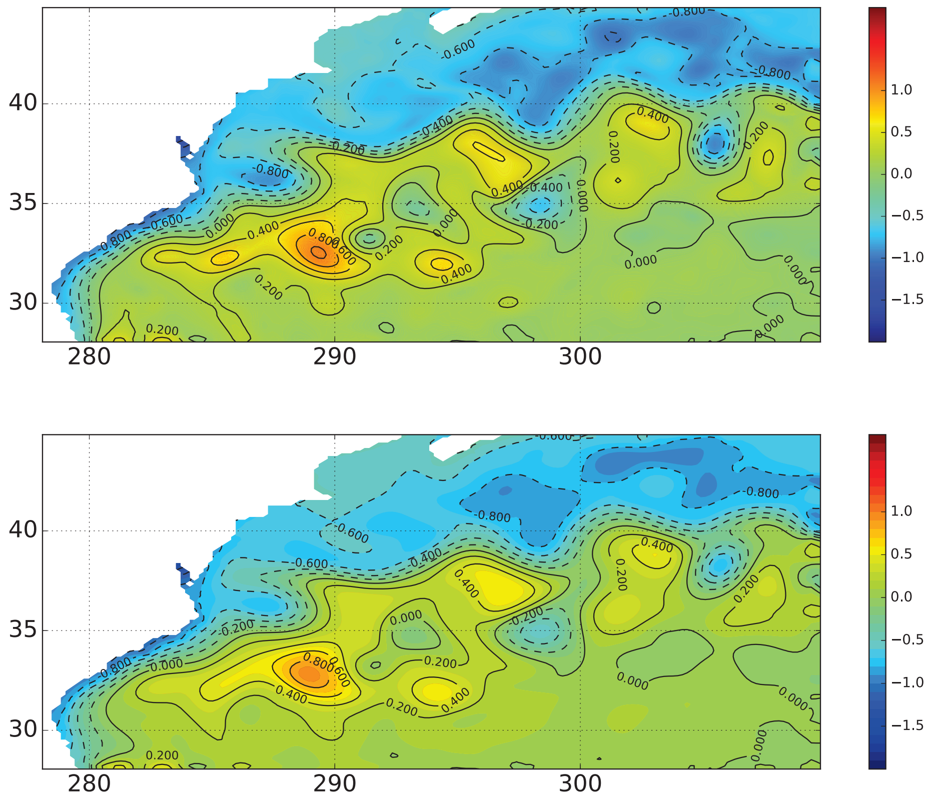


Lyapunov function versus time during 10 BFN iterations, with space-complete and time-sampled data (1 observation every 10 or 150 time steps)

SWOT observations on a QG model

Results with SWOT-like noisy data :



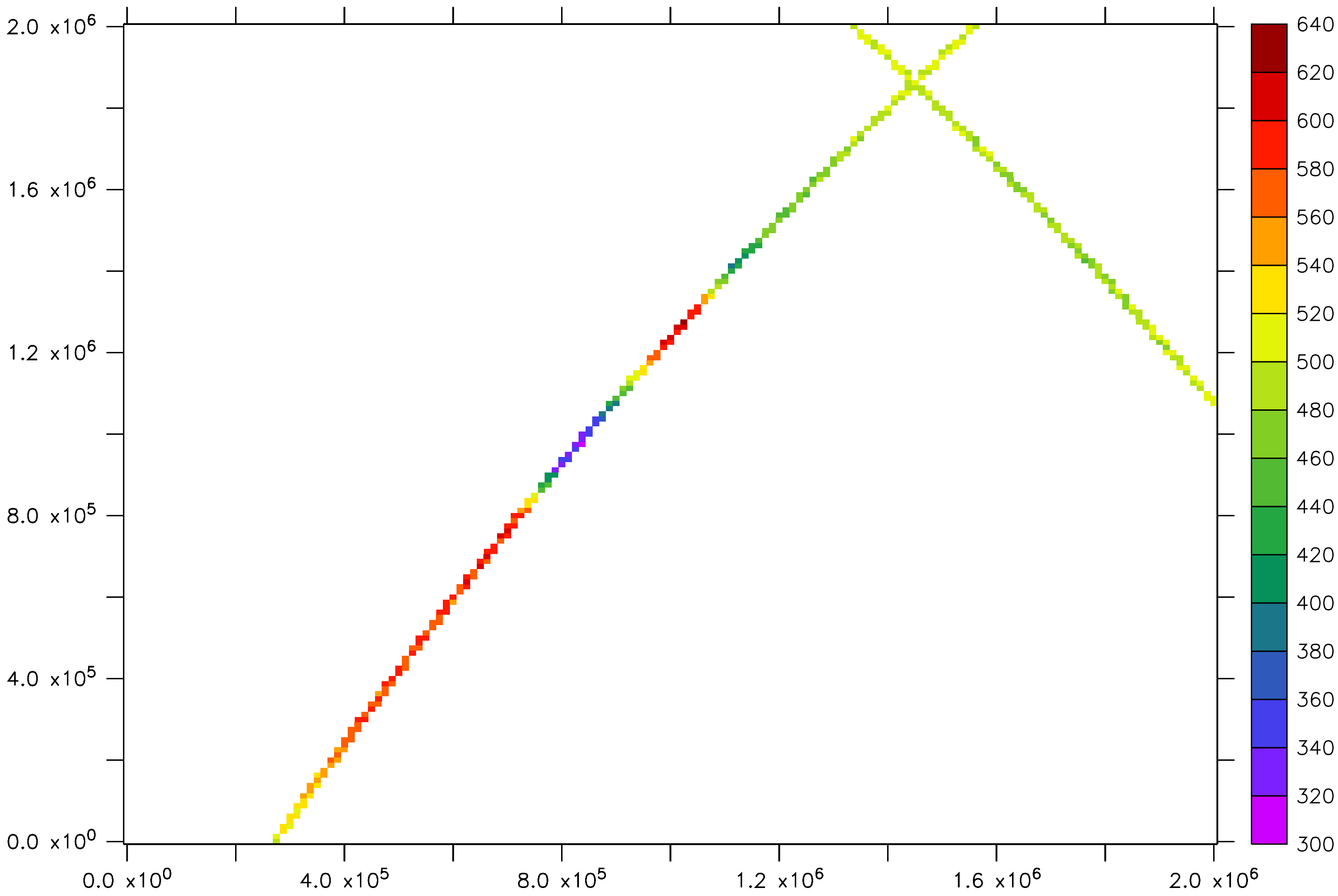


Convergence of BFN on SWOT-like noisy observations ; exact and assimilated SSH after 21 days

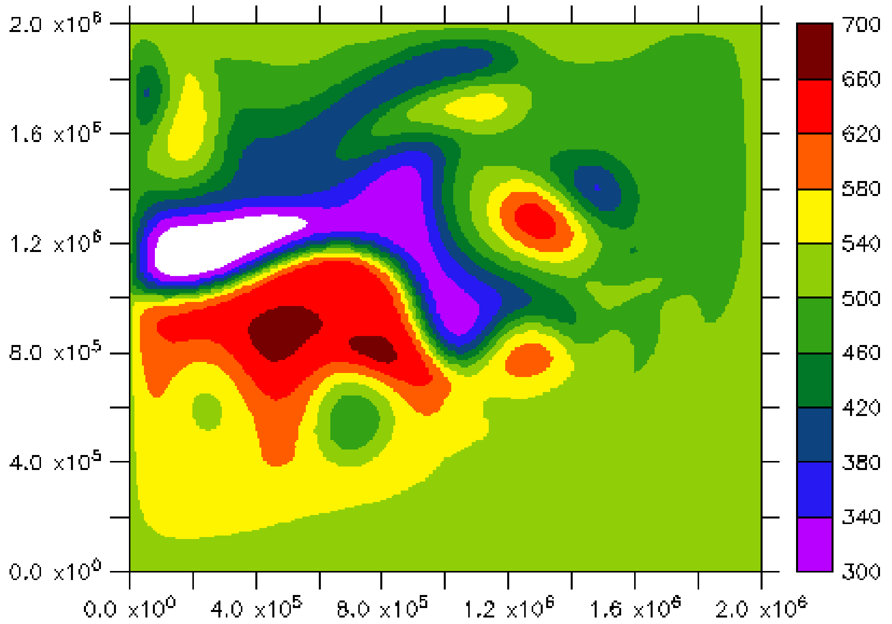
BF-reduced Kalman filter

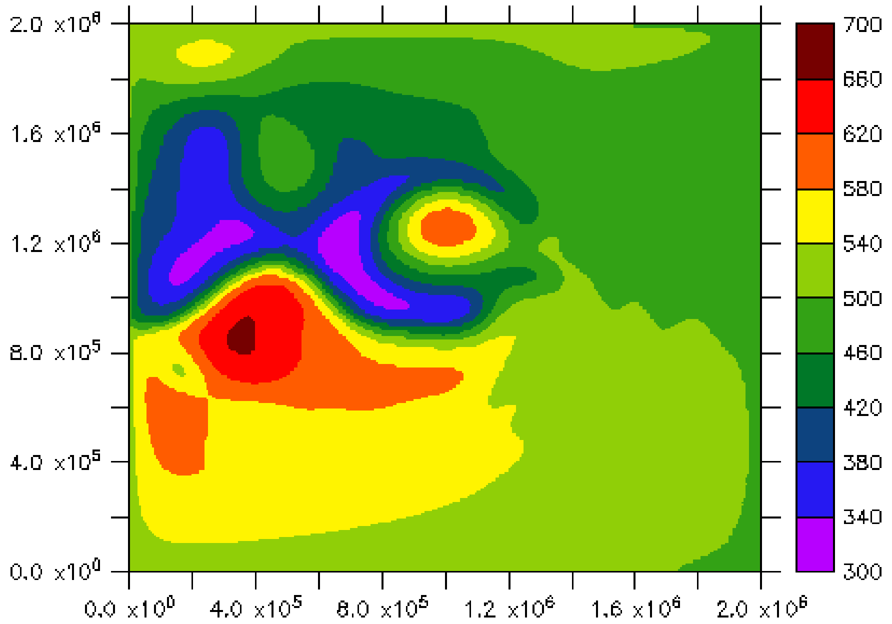
Numerical configuration :

- Shallow water model (2D), domain : 2000×2000 km
- Assimilation window : 30 days, daily assimilation
- SSH observation lines along Topex/Poseidon ground tracks
- 4 iterations of BF-SEEK : back and forth reduced Kalman filter (propagation of low rank approximations of the covariance matrices, assuming the initial ones are low rank)
- But no dynamical propagation of the errors (still in progress) ($\rightsquigarrow$ SEEK $\simeq$ optimal interpolation)



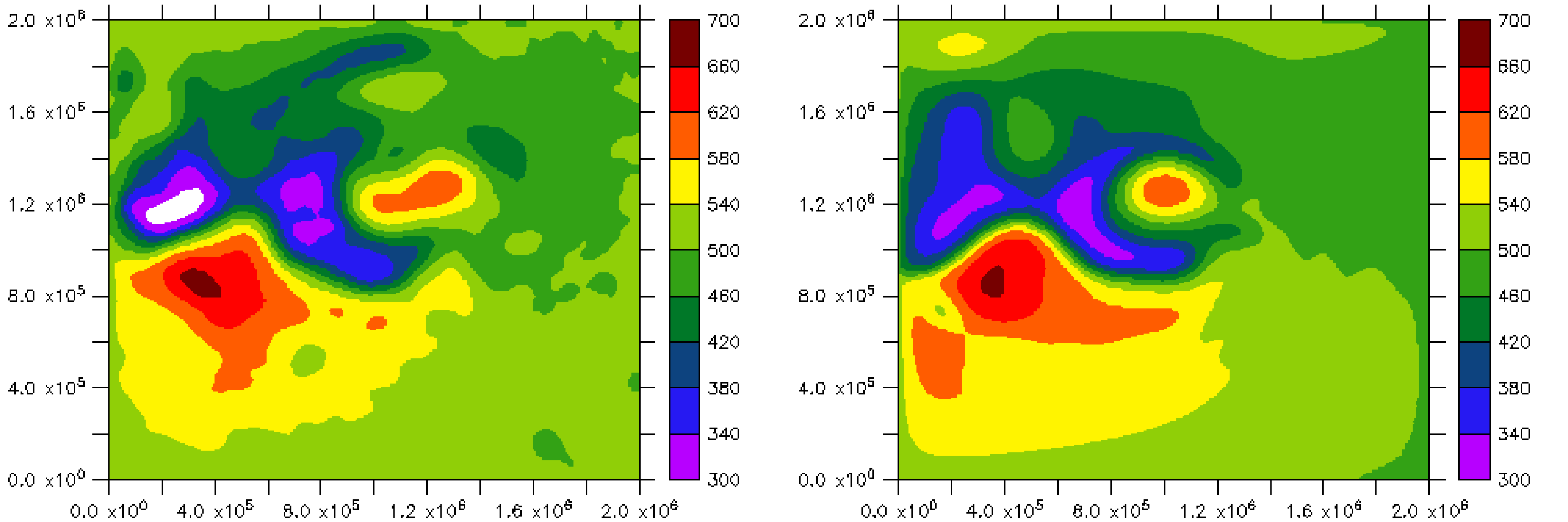
Example of assimilated data every day





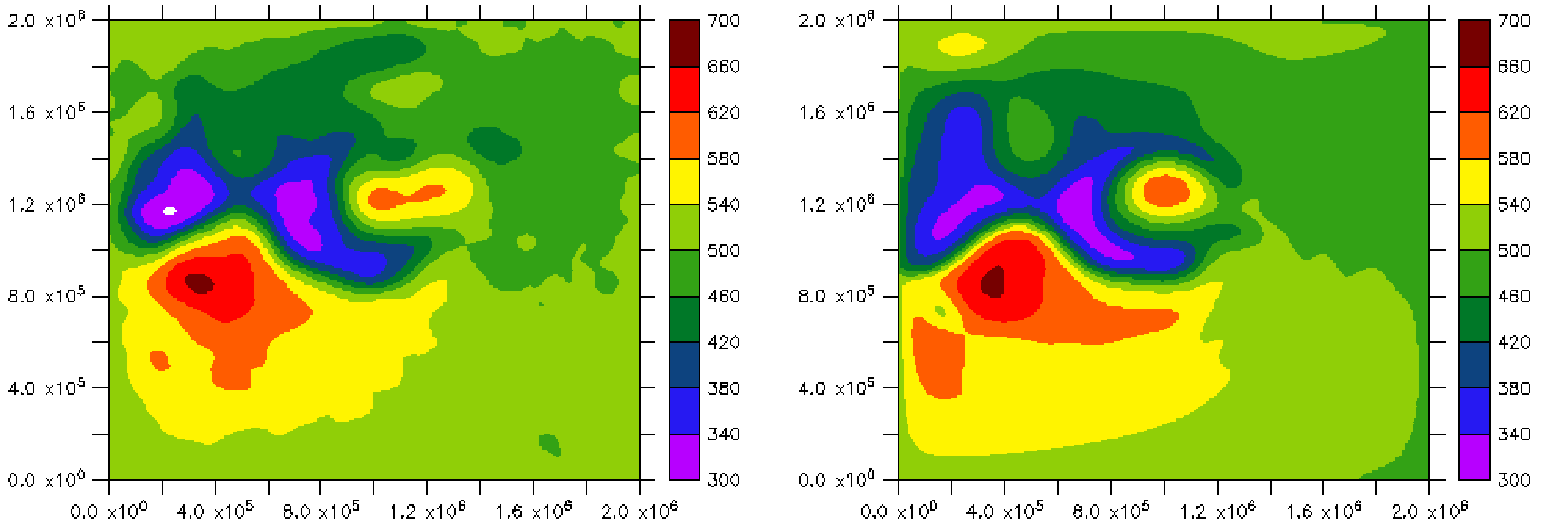
Identified SSH (left) and “true” SSH (right) after : 0 iteration of BF-SEEK

BF-SEEK



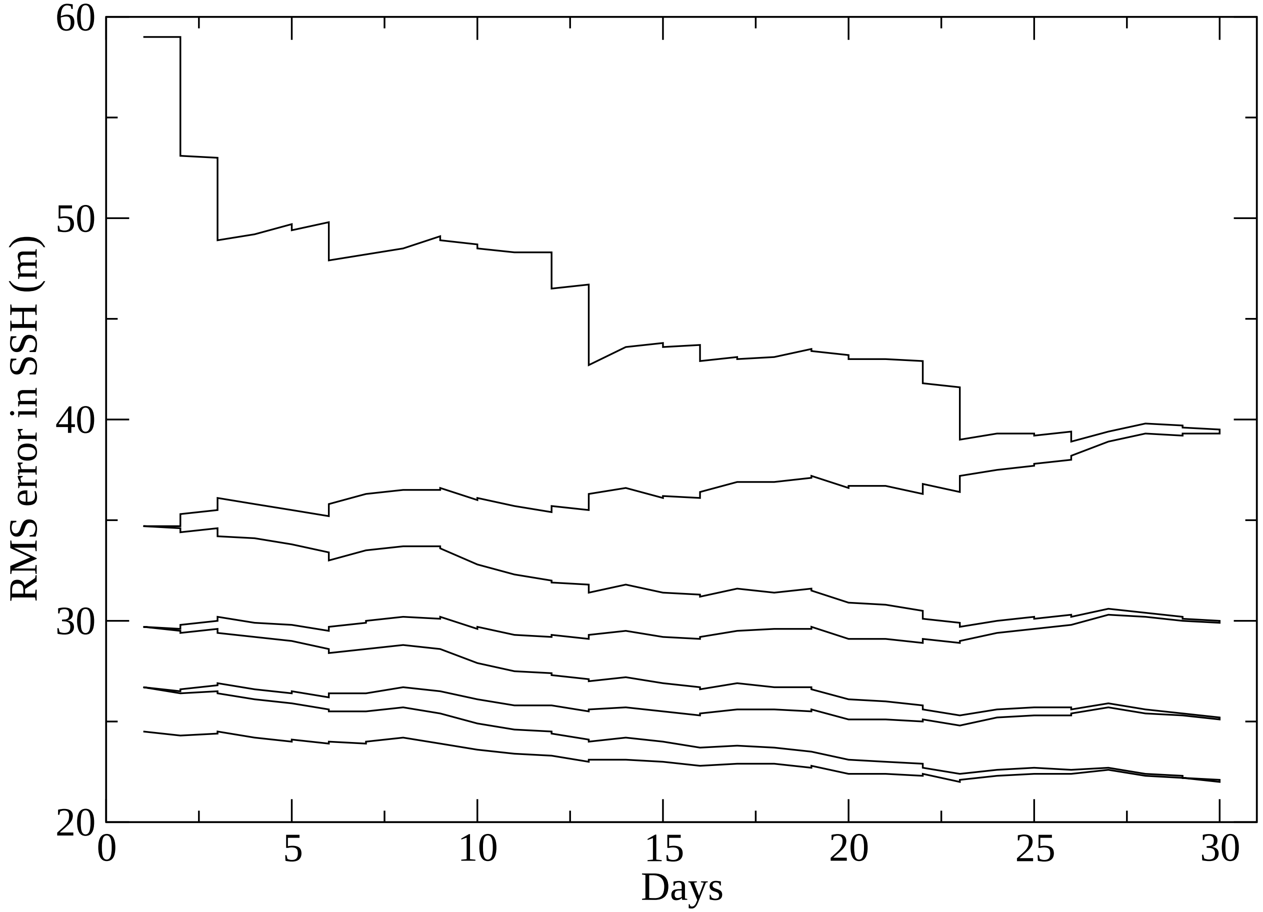
Identified SSH (left) and “true” SSH (right) after : 2 iterations of BF-SEEK

BF-SEEK



Identified SSH (left) and “true” SSH (right) after : 4 iterations of BF-SEEK

BF-SEEK



Evolution of the RMS error on the SSH during the BF-SEEK iterations.

Diffusion problem

Backward model and diffusion :

The main issue of the BFN is : how to handle diffusion processes in the backward equation ?

Let us consider only diffusion : heat equation (in 1D)

$$\partial_t u = \partial_{xx} u$$

The backward nudging model will be :

$$\partial_t \tilde{u} = \partial_{xx} \tilde{u} + K(\tilde{u} - u_{obs})$$

from time T to 0. By using a change of variable $t' = T - t$, we can rewrite the backward model as a forward one :

$$\partial_{t'} \tilde{u} = -\partial_{xx} \tilde{u} - K(\tilde{u} - u_{obs}),$$

and we can see that even if the nudging term stabilizes the model, the backward diffusion is a real issue (unbounded eigenvalues, except for discrete Laplacian).

Diffusion problem

Hopefully, in geophysical problems, diffusion is not a dominant term. The model has smoothing properties, and diffusion is small $\rightarrow$ diffusion processes are not highly unstable in backward mode, even if the model is clearly unstable without nudging.

Theoretically, there is a problem :

- **Viscous linear transport equation :** if the support of K is a strict sub-domain (i.e. some parts of the space domain are not observed), there does not exist a solution to the backward model, even in the distribution sense.
- **Viscous Burgers equation :** even if K is constant (in time and space $\Rightarrow$ full observations), the backward equation is ill-posed, as there is no stability (or continuity) with respect to the initial condition.

Without viscosity, one can prove the convergence of the BFN on these equations.

1. Nudging and observers

2. Back and Forth Nudging algorithm

⇒ 3. Diffusive BFN algorithm

4. Parameter estimation

Diffusive BFN

Diffusive free equations in the geophysical context :

In meteorology or oceanography, theoretical equations are usually diffusive free (e.g. Euler's equation for meteorological processes).

In a numerical framework, a diffusive term is added to the equations (or a diffusive scheme is used), in order to both stabilize the numerical integration of the equations, and take into consideration some subscale phenomena.

Example : weather forecast is done with Euler's equation (at least in Météo France...), which is diffusive free. Also, in quasi-geostrophic ocean models, people usually consider ∇^4 or ∇^6 for dissipation at the bottom, or for vertical mixing.

Diffusive BFN

Non viscous model - artificial diffusion term :

$$\partial_t X = F(X) + \nu \Delta X, \quad 0 < t < T,$$

where F has no diffusive terms, ν is the diffusion coefficient, and we assume that the diffusion is a standard second-order Laplacian (could be a higher order operator).

We introduce the D-BFN algorithm in this framework, for $k \geq 1$:

$$\begin{cases} \partial_t X_k = F(X_k) + \nu \Delta X_k + K(\mathcal{Y} - H(X_k)), \\ X_k(0) = \tilde{X}_{k-1}(0), \quad 0 < t < T, \end{cases}$$

$$\begin{cases} \partial_t \tilde{X}_k = F(\tilde{X}_k) - \nu \Delta \tilde{X}_k - K'(\mathcal{Y} - H(\tilde{X}_k)), \\ \tilde{X}_k(T) = X_k(T), \quad T > t > 0. \end{cases}$$

This backward equation can be rewritten in forward mode ($t' = T - t$) :

$$\partial_{t'} \tilde{X}_k = -F(\tilde{X}_k) + \nu \Delta \tilde{X}_k + K'(\mathcal{Y} - H(\tilde{X}_k)), \quad \tilde{X}_k(t' = 0) = X_k(T),$$

which can easily be solved (only the physical model has an opposite sign).

Linear transport - smoothing equation

If we apply the D-BFN algorithm to a linear transport equation (model F) :

$$\partial_t u + a(x)\,\partial_x u = 0, \quad t \in [0,T],\ x \in \Omega, \quad u(t=0) = u_0,$$

that we will solve numerically with a small diffusion term (for stability and subscale modelling), then the D-BFN algorithm converges. At the limit $k \to \infty$, u_k and $\tilde{u}_k$ tend to $u_\infty(x)$ solution of

$$\nu \partial_{xx} u_\infty + K(u^0_{obs}(x) - u_\infty) = 0,$$

or equivalently

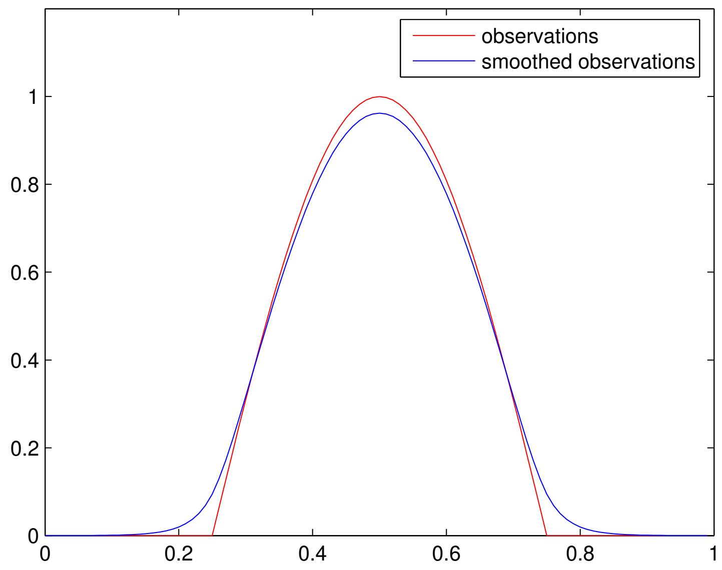
$$-\frac{\nu}{K}\partial_{xx} u_\infty + u_\infty = u^0_{obs}.$$

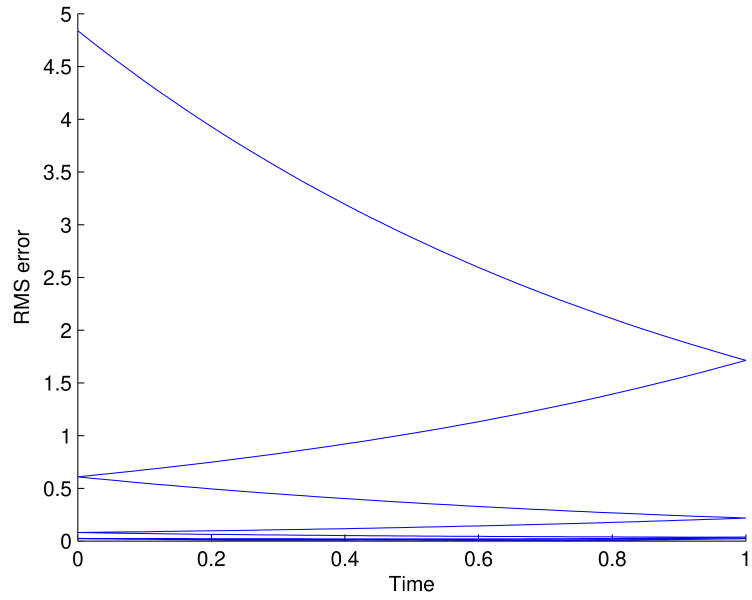
This equations is well known in signal or image processing, as being the standard linear diffusion restoration equation. In some sense, u_∞ is the result of a smoothing process on the observations u_{obs}, where the degree of smoothness is given by the ratio $\frac{\nu}{K}$.

Linear transport - DBFN

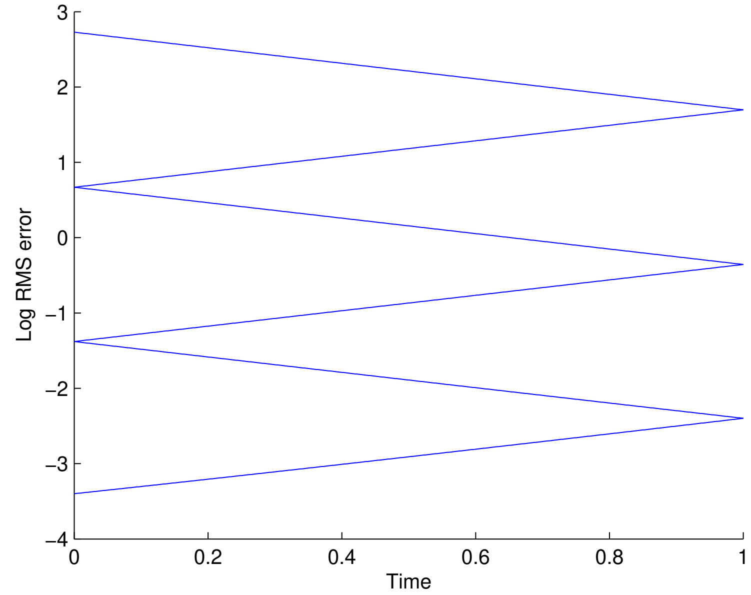
(DBFN on linear transport)

Linear transport - DBFN





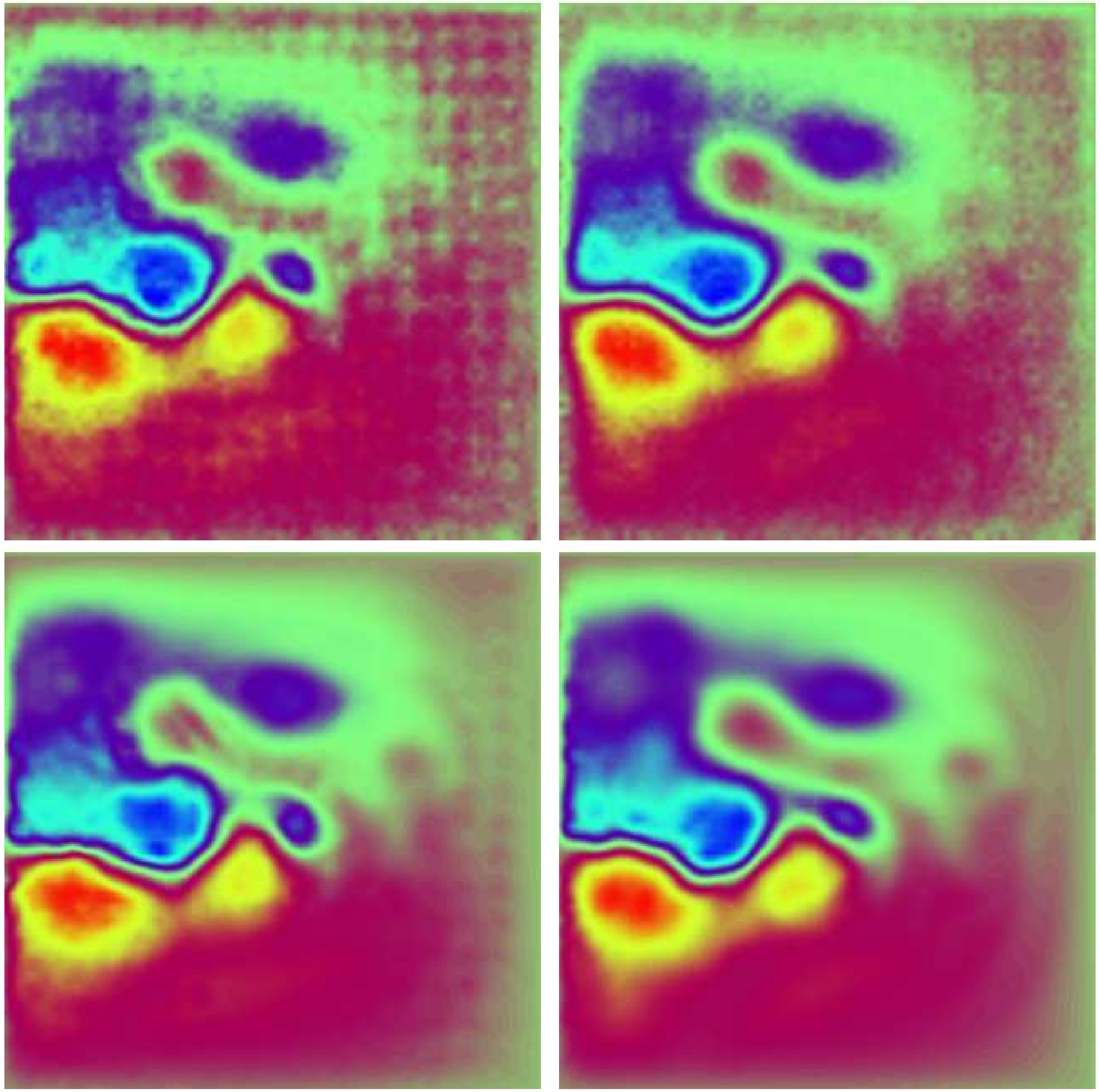
Initial condition of the observation and corresponding smoothed solution ; RMS difference between the BFN iterates and the smoothed observations ; same in semi-log scale.



Shallow water model

BFN (5 iter.)

DBFN (5 iter.)

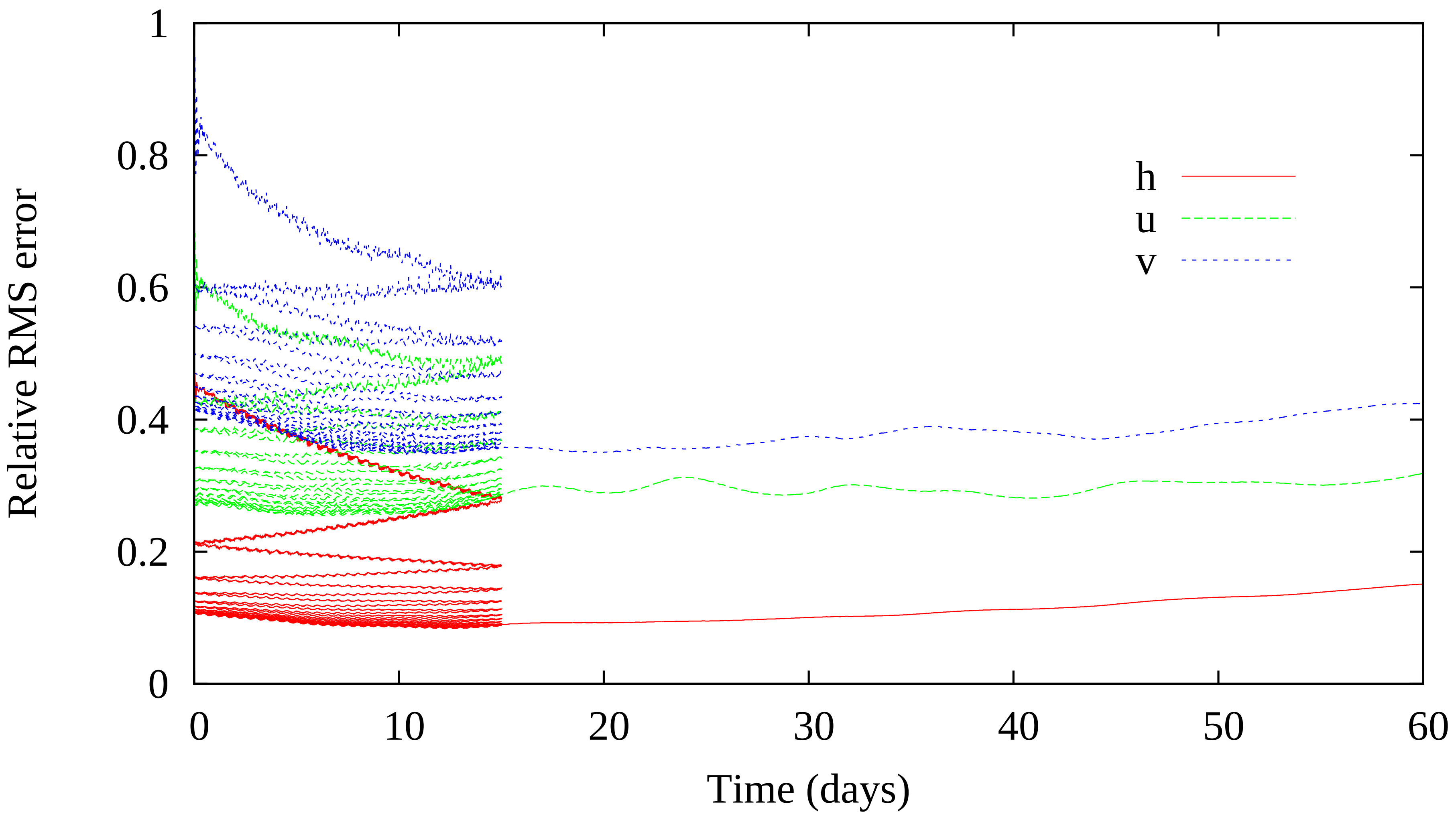


4DVAR (50 iter.)

True state

Initial condition (sea surface height) identified by : BFN (5 iterations, converged), DBFN (5 iterations, converged), 4DVAR (50 iterations, converged), true solution.

Shallow water model



Evolution of relative RMS error of the DBFN solution during the iterations (15 first days, corresponding to the assimilation window), and the forecast (from day 15 to day 60, corresponding to the forecast window).

Full primitive ocean model

Primitive equations : Navier-Stokes equations (velocity-pressure), coupled with two active tracers (temperature and salinity).

Momentum balance :

$$\frac{\partial U_h}{\partial t} = -\left[(\nabla \wedge U) \wedge U + \frac{1}{2} \nabla (|U|^2) \right]_h - f.z \wedge U_h - \frac{1}{\rho_0} \nabla_h p + D^U + F^U$$

Incompressibility equation : $\nabla . U = 0$

Hydrostatic equilibrium : $\dfrac{\partial p}{\partial z} = -\rho g$

Heat and salt conservation equations :

$$\frac{\partial T}{\partial t} = -\nabla.(TU) + D^T + F^T \quad \text{(+ same for S)}$$

Equation of state : $\rho = \rho(T, S, p)$

Full primitive ocean model

Free surface formulation : the height of the sea surface η is given by

$$\frac{\partial \eta}{\partial t} = -div_h((H+\eta)\bar{U_h}) + [P - E]$$

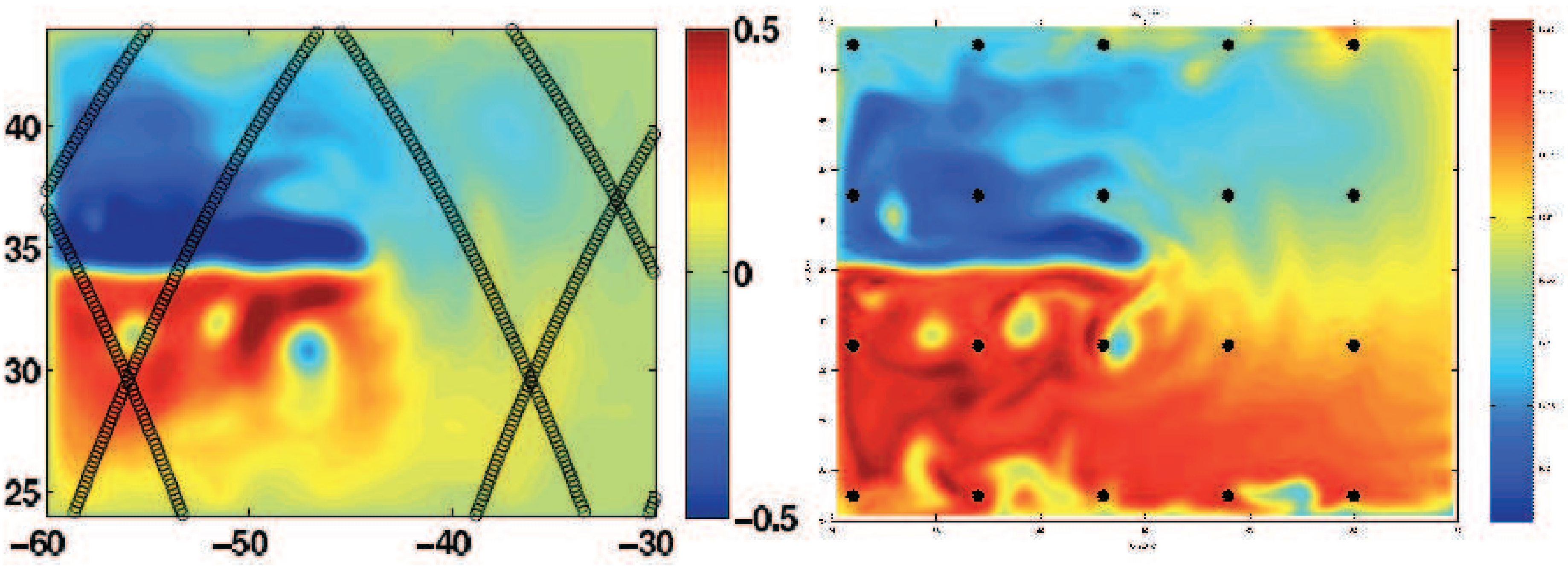
The surface pressure is given by : $p_s = \rho g \eta.$

This boundary condition is then used for integrating the hydrostatic equilibrium and calculating the pressure.

Numerical experiments : double gyre circulation confined between closed boundaries (similar to the shallow water model). The circulation is forced by a sinusoidal (with latitude) zonal wind.

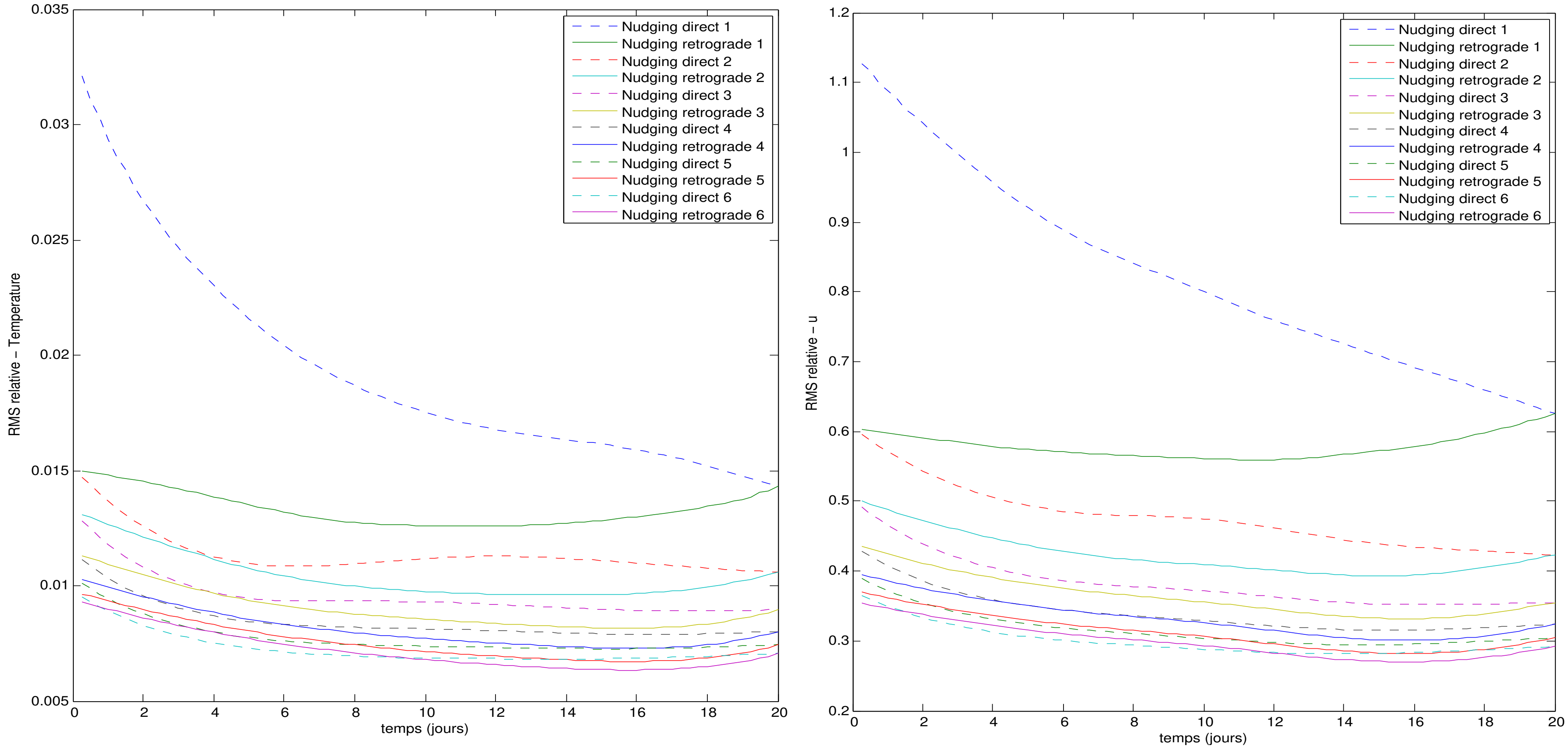
Twin experiments : observations are extracted from a reference run, according to networks of realistic density : SSH is observed similarly to TOPEX/POSEIDON, and temperature is observed on a regular grid that mimics the ARGO network density.

Full primitive ocean model



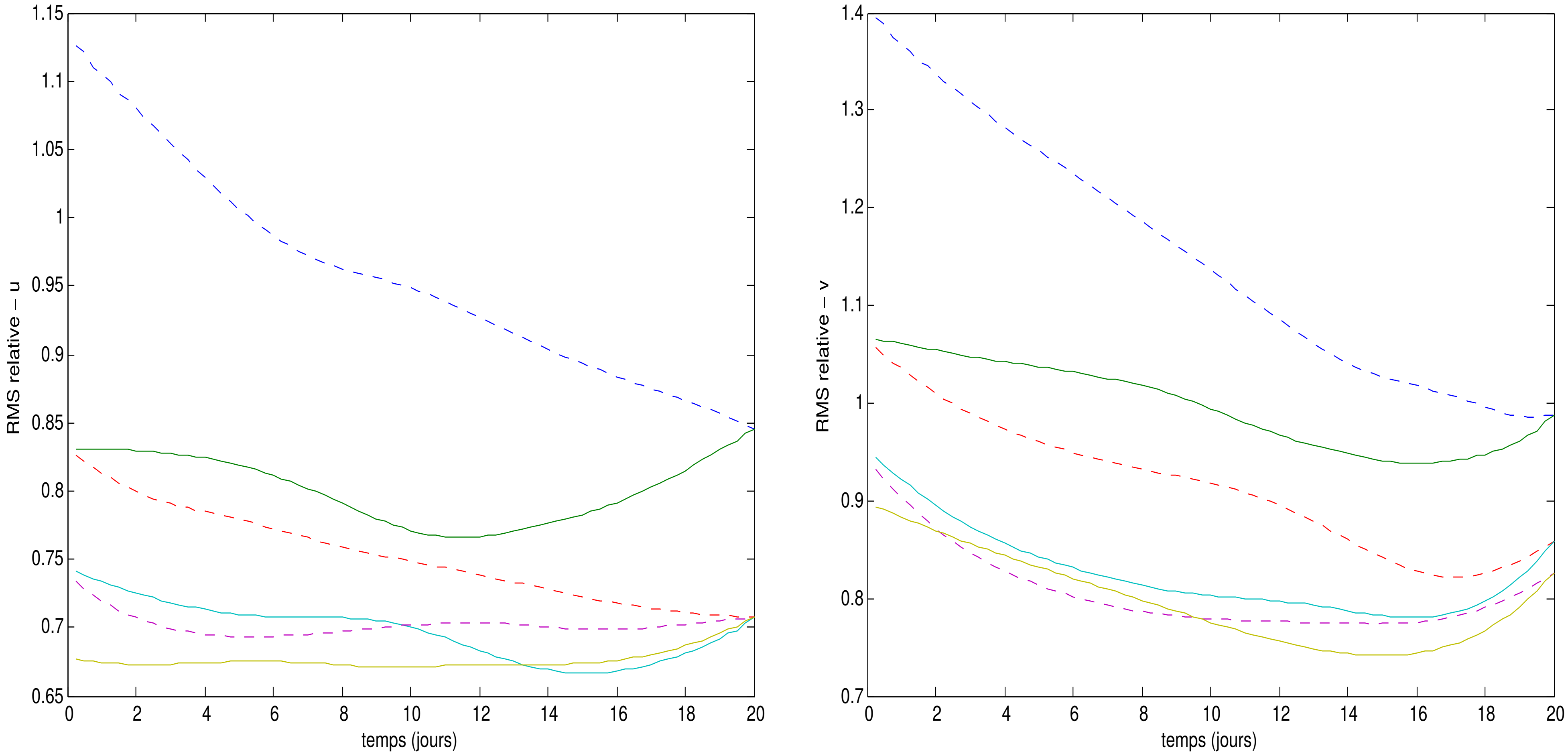
Example of observation network used in the assimilation : along-track altimetric observations (Topex-Poseidon) of the SSH every 10 days ; vertical profiles of temperature (ARGO float network) every 18 days.

Numerical results



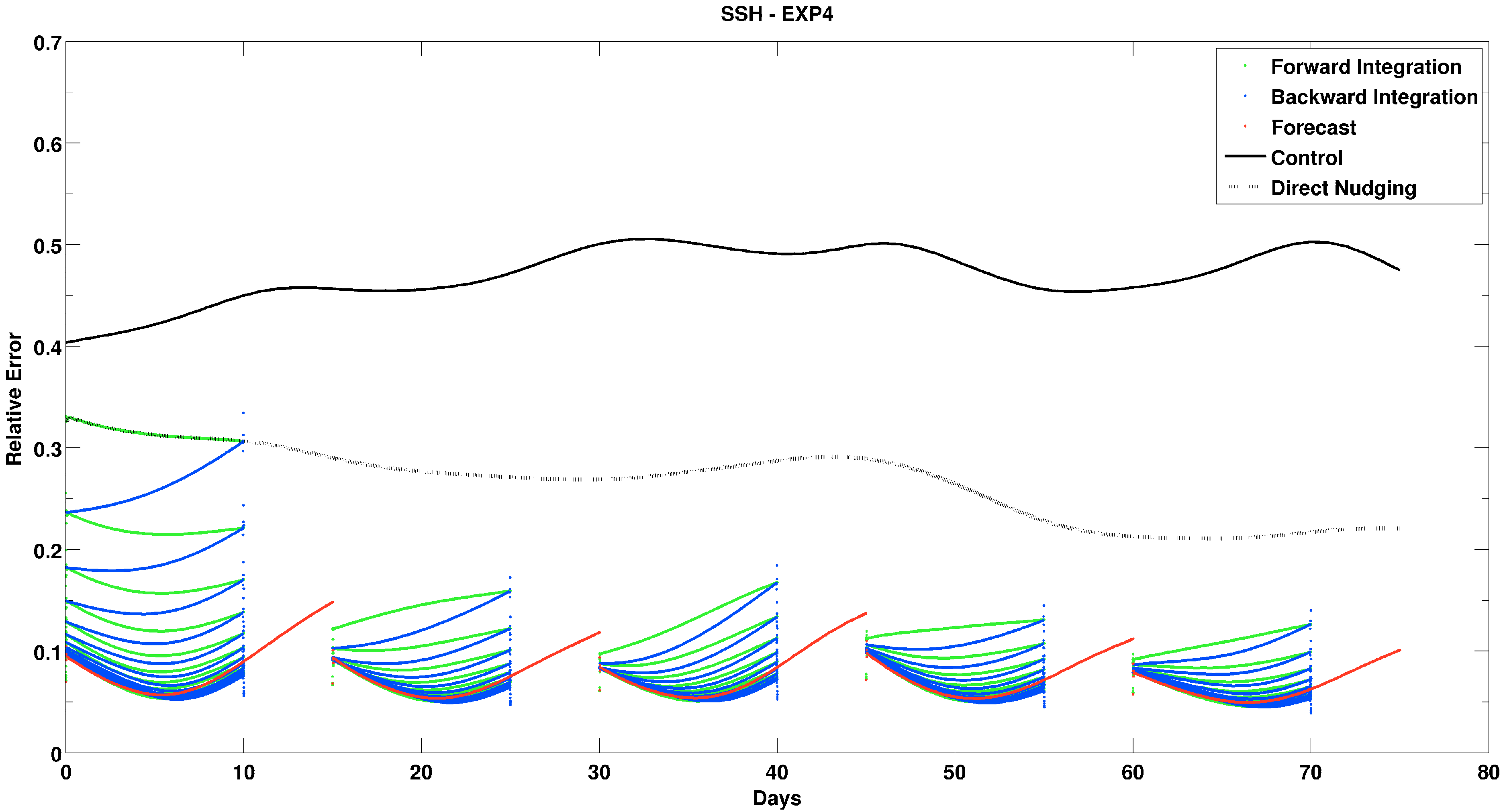
Relative RMS error of the temperature (left) and longitudinal velocity (right), 6 iterations of D-BFN (nudging terms in the temperature and SSH equations only), with full and unnoisy SSH observations every day.

Numerical results



Relative RMS error of the longitudinal and transversal velocities, 3 iterations of D-BFN (nudging terms in the temperature and SSH equations only), with "realistic" SSH observations (T/P track + 15% noise).

Numerical results



Evolution of the errors during the Back and Forth iterations and during the forecast phase. In black : evolution of the error for the control and direct nudging experiments.

1. Nudging and observers

2. Back and Forth Nudging algorithm

3. Diffusive BFN algorithm

⇒ 4. Parameter estimation

P-BFN : parameter estimation

Let assume now that the parameter $a(x)$ of the transport equation is unknown. We want to estimate both the model state u and parameter a. We add an ad hoc equation for the time independent parameter :

$$\begin{cases} \partial_t u(t,x) + a(x)\partial_x u(t,x) = 0, & u(0,x) = u_0(x), \\ \partial_t a(t,x) = 0, & a(0,x) = a(x). \end{cases}$$

Then we apply the BFN algorithm to this coupled system, and we add feedback terms to both equations, using only observations on the state u :

$$\begin{cases} \partial_t \hat{u}(t,x) + \hat{a}(t,x)\partial_x \hat{u}(t,x) = K_u(u_{obs}(t,x) - \hat{u}(t,x)), \\ \partial_t \hat{a}(t,x) = K_a \mathcal{F}(u_{obs}(t,x) - \hat{u}(t,x)), \end{cases}$$

where $\mathcal{F}$ is a feedback function involving spatial differential operators, such that there exists a Lyapunov function that decreases in time.

Then, we can prove that both u and a can be reconstructed.

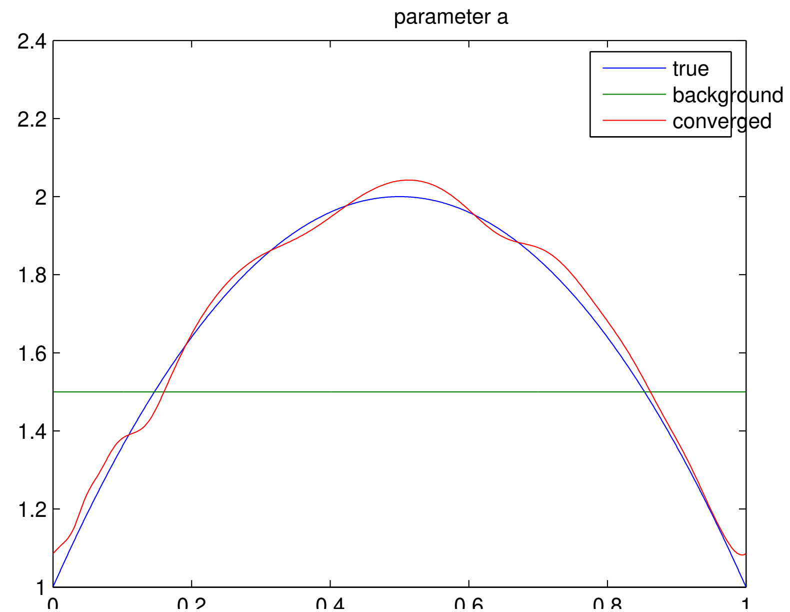
P-BFN - Transport equation

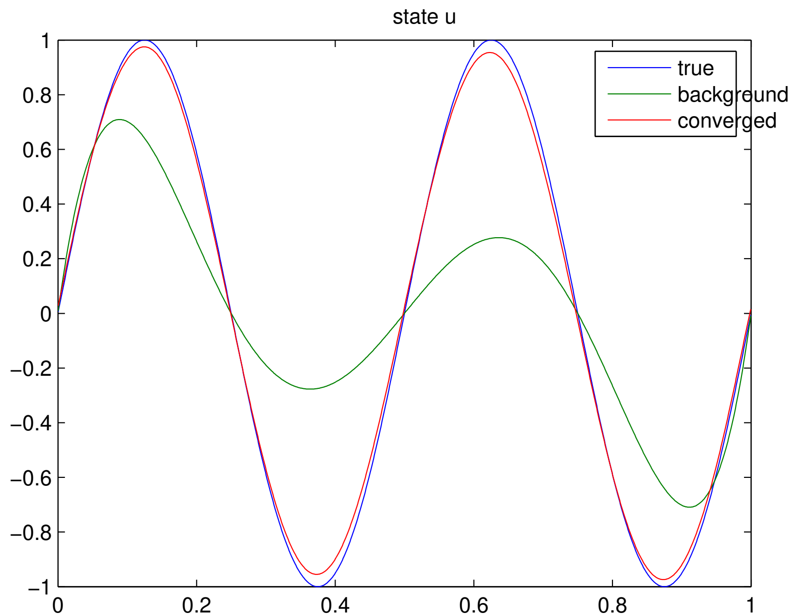
State

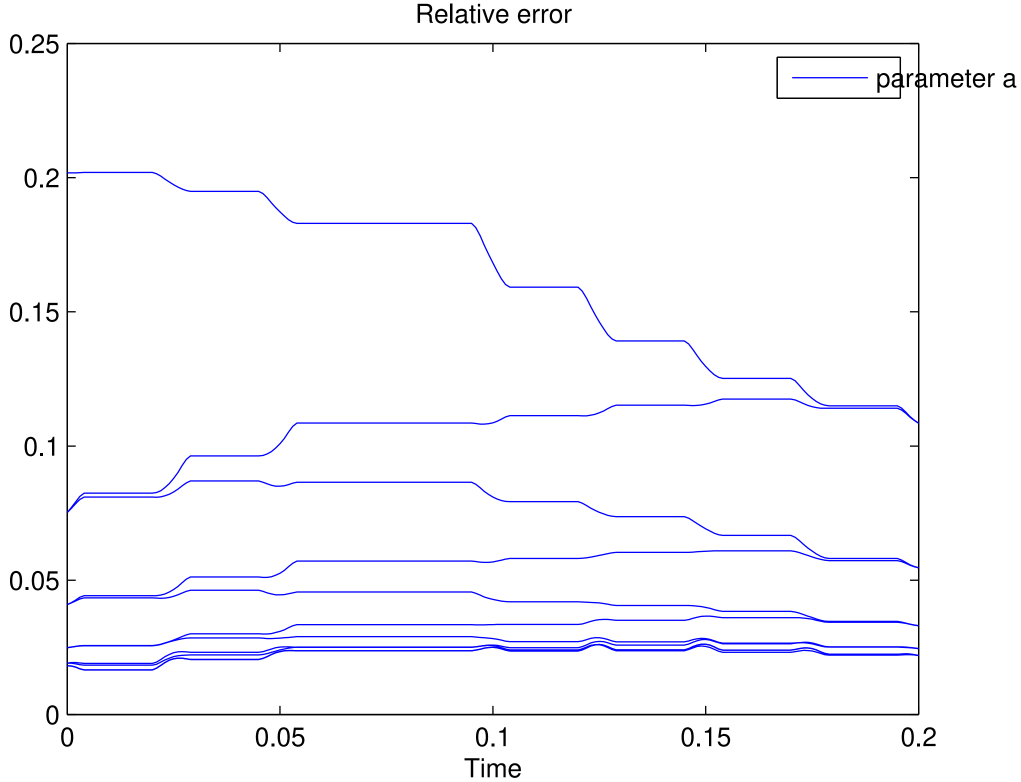
Parameter

(P-BFN on linear transport)

P-BFN - Transport equation







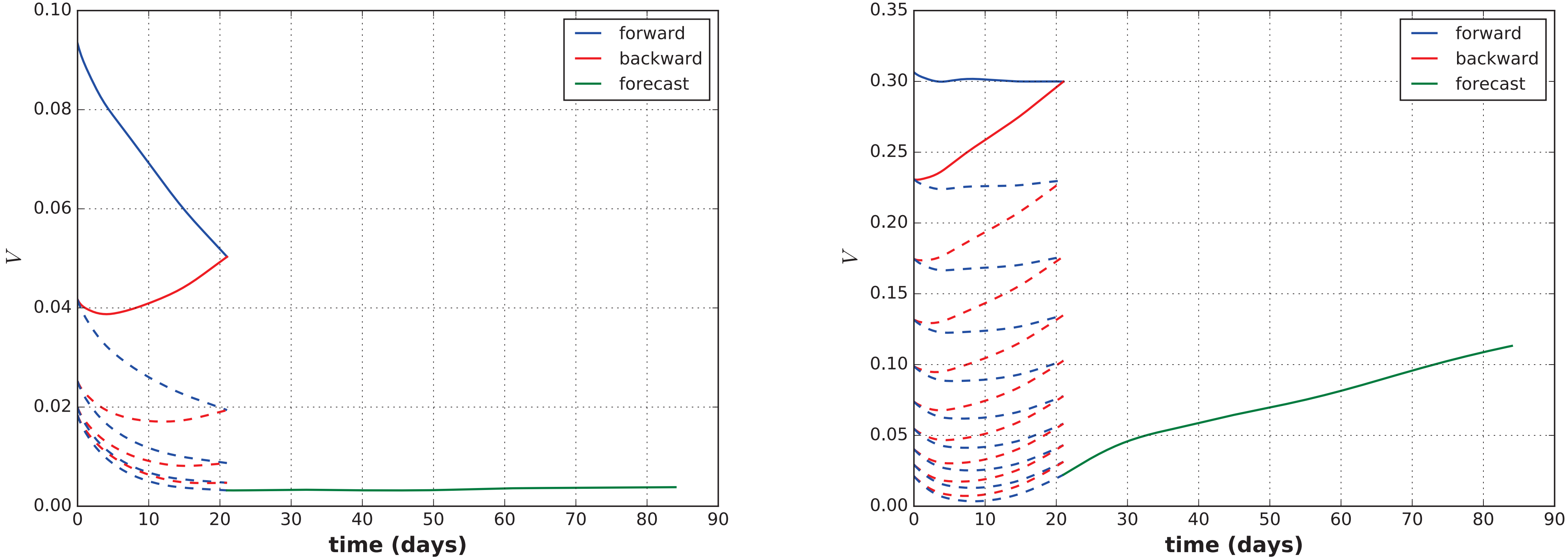
Comparisons between P-BFN identified solution and true solution (left)

Error vs P-BFN iterations (right)

P-BFN - QG model with SWOT data

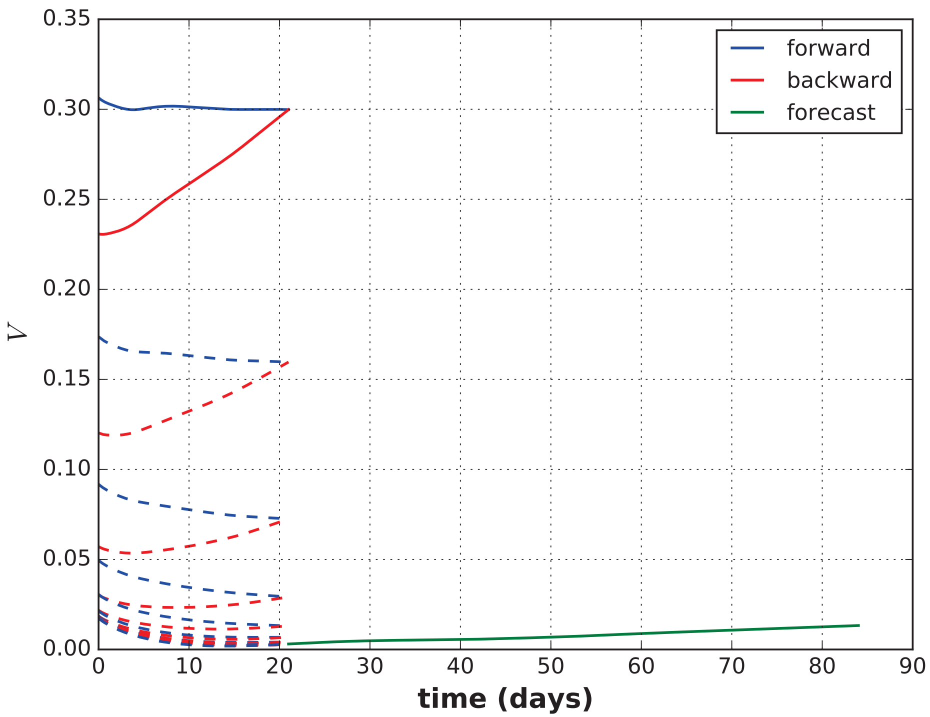
Back to QG model, with SWOT data

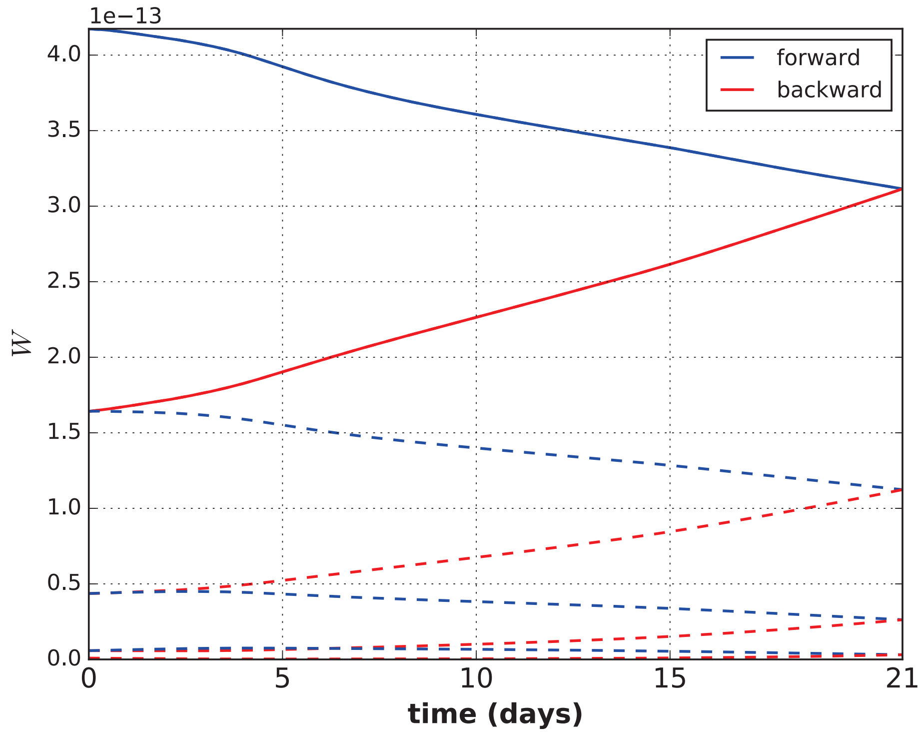
Parameter : phase speed c (or barotropic deformation wavenumber $\kappa = \frac{f^2}{c^2}$)



SSH error with : true parameter, wrong parameter.

P-BFN - QG model with SWOT data





SSH error with a jointly estimated parameter ; Parameter error.

Conclusions

Back and Forth Nudging algorithm :

- Easy implementation (no linearization, no adjoint state, no minimization process)
- Very efficient in the first iterations (faster convergence)
- Lower computational and memory costs than other DA methods
- Stabilization of the backward model
- Excellent preconditioner for 4D-VAR (or Kalman filters)

Diffusive BFN algorithm :

- Converges even faster, with smaller backward nudging coefficients
- Still produces very precise forecasts

Parameter estimation BFN algorithm :

- Efficient estimation of model parameters at almost no additional computational cost
- But the feedback term on the parameter is equation dependent $\rightsquigarrow$ increase of the human brain cost

Perspectives

Extension to more (but not too) complex Back and Forth Observers :

- Observers for a coupled system (image + compresible Navier-Stokes) : reconstruction of density and velocity fields from image measurements
- Use of physical considerations : e.g. geostrophic equilibrium (Coriolis force $\simeq$ pressure gradient) to correct non observed variables

Extension to parameter estimation :

- Add an equation for the parameter (e.g. $\frac{d\alpha}{dt} = 0$), observe the physical variables, and try to build an observer that corrects all variables (including the parameter)
- Use of observers in a similar way as Kalman filtering for parameter estimation ($\rightsquigarrow$ Fourier decompositions, energy estimates, Lyapunov theory, ...)

THANK YOU FOR YOUR ATTENTION !