Taking into account correlated observation errors by progressive assimilation of multiscale information

Vincent Chabot, Maëlle Nodet, Arthur Vidard

Taking into account correlated observation errors by progressive assimilation of multiscale information

Vincent Chabot, Maëlle Nodet, Arthur Vidard — Univ. Grenoble Alpes, INRIA, France

Abstract

- The description of **correlated observation error statistics** is a challenge in data assimilation.
- Currently, the observation errors are assumed uncorrelated (the covariance matrix is diagonal) which is a **severe approximation** that leads to suboptimal results.
- It is possible to use **multi-scale transformations** to retain the diagonal matrix approximation while accounting for some correlation. However this approach can lead to some convergence problems due to scale interactions.
- We propose an **online scale selection algorithm** that improves the convergence properties in such case.

Motivation: account for correlated satellite observation errors

- Numerical weather prediction requires the determination of the initial state of the system. Indeed, the true state, at a given moment and in all points of space, is not accessible. In order to retrieve an optimal initial condition one uses the so called data assimilation methods that combine information from observations, model equations and their respective error statistics.
- Since the late 70s, satellites are a dominant source of information. Errors associated to such data are **highly correlated in space**, which can be detrimental if this is not properly accounted for. However their density in space allows for the **efficient use of multi-scale transformation**, which in turn permit a cheap but good approximation of said error statistics representation.
- For homogeneous spatially correlated Gaussian observation errors this approach is very efficient.
- For more complex errors, however, it can severely damage the convergence properties of the assimilation methods.
- Here we present, through a simple case mimicking a laboratory experiment, an illustration of the above-mentioned problem and a possible solution using scale selection during the assimilation process.

Data assimilation: Principle

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
- errors statistics

What is data assimilation for?

- historically: initial state estimation, for weather forecasting.
- today, many other applications...
- ... and many other application domains.

Best Linear Unbiased Estimation (BLUE)

Least squares analysis: 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:

- $\mathbf{y}^o$ incomplete observations, errors ϵ^o unbiased, covariance matrix $\mathbf{R}$
- $\mathbf{x}^f = \mathbf{x}^t + \epsilon^f$, ϵ^f background errors unbiased, covariance $\mathbf{P}^f$
- observation operator $\mathcal{H} = \mathbf{H}$ maps linearly the input parameters to the observations

Solution: Best linear unbiased estimator (BLUE):

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

Variational equivalence: $\mathbf{x}^a$ is the minimizer of

$$J(\mathbf{x}) = (\mathbf{x} - \mathbf{x}^f)^T\mathbf{P}^f(\mathbf{x} - \mathbf{x}^f) + (\mathbf{H}\mathbf{x} - \mathbf{y}^o)^T\mathbf{R}(\mathbf{H}\mathbf{x} - \mathbf{y}^o)$$

Variational Data Assimilation

Let $\mathcal{M}$ be a dynamical model describing the evolution of the state variable X in space and time:

$$\left\| \begin{array}{l} \partial_t X(X_0, x, t) + \mathcal{M}(X(X_0, x, t)) = 0 \\ X(X_0, x, t_0) = X_0 \end{array} \right.$$

Let $Y(t)$ be (partial) observations of this state variable.
The aim of data assimilation is to estimate an optimal initial condition X_0^a (often called analysed state) so that it is not far from the first guess X_0^b (in general coming from a previous forecast), and that the model trajectory $X(X_0^a, x, t)$ is close to the observations $Y(t)$.

This is done by defining X_0^a as the minimum of the cost function:

$$\begin{aligned} J(X_0) &= J_b(X_0) + J_o(X_0) \\ &= \frac{1}{2}\|X_0 - X_0^b\|_{\mathbf{V}}^2 + \\ &\quad \frac{1}{2}\sum_{t_i=t_0}^{t_f} ||Y(t_i) - \mathcal{H}(X(X_0, x, t_i))||_{\mathcal{O}}^2 \end{aligned}$$

where $\mathbf{V}$ is the model state space, $\mathcal{O}$ the observation space and $\mathcal{H} : \mathbf{V} \mapsto \mathcal{O}$ the observation operator. Usually, in variational data assimilation, the minimisation is done using a gradient descent type algorithm and the gradient is computed using adjoint methods.
Typically in data assimilation one uses the Mahalanobis distance:

$$\|.\|_{\mathbf{V}}^2 = \|.\|_{\mathbf{B}}^2, \quad \|.\|_{\mathcal{O}}^2 = \|.\|_{\mathbf{R}}^2, \quad \text{with } \|x\|_{\mathbf{K}}^2 = x^T\mathbf{K}^{-1}x$$

where $\mathbf{R}$ and $\mathbf{B}$ are the observation and background error covariance matrices respectively.

Error covariance matrix representation

- Crucial choice for $\mathbf{B}$ and $\mathbf{R}$: drive the way information is spread and how redundancy of error is delt with.
- So far: strong research effort for $\mathbf{B}$ modelling.
- $\mathbf{R}$ matrix: mostly assumed diagonal.
  - Main reason for this: management greatly simplified.
  - Number of observations in general non constant over time $\rightarrow$ prevents from using the same $\mathbf{R}$ matrix.
  - Consequelntly: at each assimilation cycle, a new $\mathbf{R}$ matrix should be formed and inverted.

Our work: change of variable, which allows diagonal $\mathbf{R}$ + non trivial correlations.

- Linear changes of variable $\mathbf{A}$: assuming observation errors to be additive, unbiased and gaussian, i.e. $Y = Y^t + \epsilon$ with $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$, Y^t being the true signal, then $\mathbf{A}Y = \mathbf{A}Y^t + \beta$ with $\beta \sim \mathcal{N}(\mathbf{0}, \mathbf{ARA}^T)$.
- Choice of $\mathbf{A}$ such that $\mathbf{D_A} = \text{diag}(\mathbf{ARA}^T) \simeq \mathbf{ARA}^T \rightarrow$ we retain a diagonal approximation:

$$\begin{aligned} &(Y - \mathcal{H}(X))^T\mathbf{R}^{-1}(Y - \mathcal{H}(X)) \\ &\quad \simeq (Y - \mathcal{H}(X))^T\mathbf{A}^T\mathbf{D_A}^{-1}\mathbf{A}(Y - \mathcal{H}(X)) \end{aligned}$$

- Here: orthonormal wavelet transform for the operator $\mathbf{A}$ (see more options in 2015 reference).

Experimental settings

Experimental framework: mimics the drift of a vortex on a turntable. CORIOLIS experimental turntable (Grenoble, France): simulates the evolution of a vortex in the atmosphere. 1 rotation of the tank = 1 earth rotation.
Vortex creation: by stirring the water and made visible thanks to the addition of a passive tracer (fluorescein). Photographs of the vortex constitute the observed image sequence.

Numerical configuration
Shallow-water equations for the horizontal velocity $w(x, t) = (u(x, t), v(x, t))$, and the water elevation $h(x, t)$:

$$\begin{cases} \partial_t u - (f + \zeta)v + \partial_x B &= -ru + \kappa\Delta u \\ \partial_t v + (f + \zeta)u + \partial_y B &= -rv + \kappa\Delta v \\ \partial_t h + \partial_x(hu) + \partial_y(hv) &= 0. \end{cases}$$

with ζ the relative vorticity and B the Bernoulli potential:

$$\zeta = \partial_x v - \partial_y u, \quad B = g^* h + \frac{u^2 + v^2}{2}$$

where g^* is the reduced gravity. Coriolis parameter: β-plane, $f = f_0 + \beta y$, κ diffusion coeff., r bottom friction coeff.

Observation operator
Vortex temporal evolution: fluorescein concentration evolution, observed by an image sequence of the concentration of a passive tracer q transported by the velocity field w:

$$\begin{cases} \partial_t q + \nabla q \cdot w - \nu_T\Delta q = 0 \\ q(t_0) = q_0. \end{cases}$$

so that:

$$\mathcal{H}(X_{t_i}) = q(t_i).$$

Twin Experiments Configuration
Synthetic observations: created thanks to a model simulation from a known "true state"; then an assimilation experiment is performed starting from another "background" state using the synthetic observations. Result can be compared with the synthetic truth.

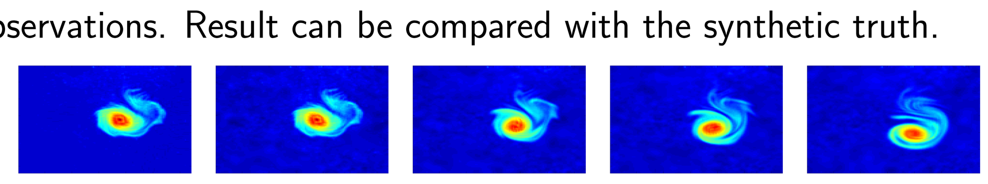

Figure caption: "True" initial concentration of the passive tracer (first left) and noisy observations at initial time, after 90 minutes, 150 min and 270 min (right).

Results 1: Homogeneous observation error

The observation are obtained by adding a spatially correlated Gaussian noise to selected snapshot of the true trajectory

$$Y_{t_i} = Y_{t_i}^t + \epsilon \quad \text{with } \epsilon = \sim \mathcal{N}(\mathbf{R}, \mathbf{I})$$

In order to mimic the usual approach, only diagonal approximations of $\mathbf{R}$ are used:

$$\mathbf{R}_{pix} = \text{diag}(\mathbf{R}) \qquad \mathbf{R}_{wav} = \mathbf{D_A} = \text{diag}(\mathbf{ARA}^T)$$

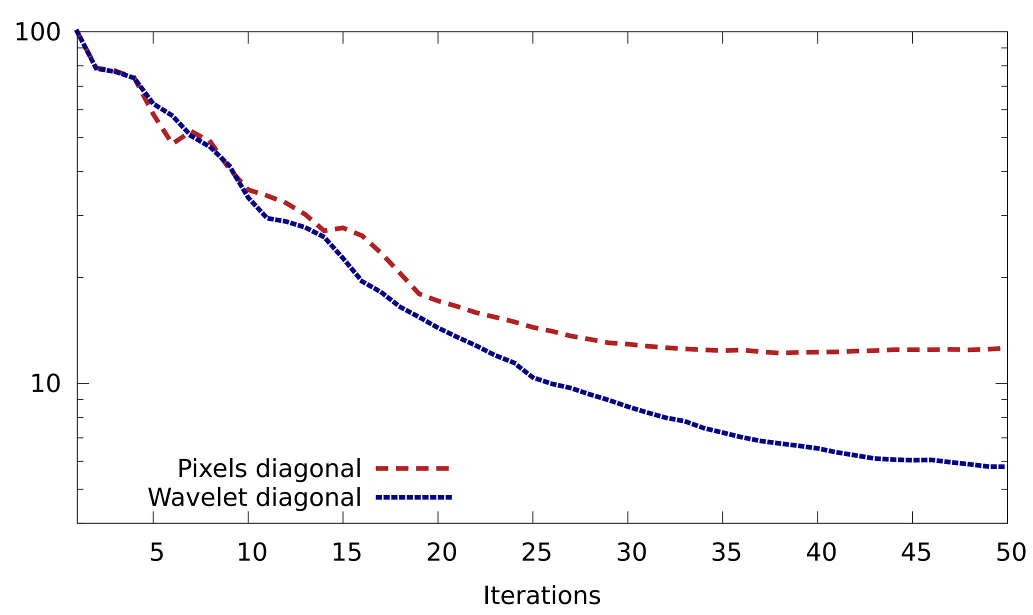

Figure caption: Ratio of residual errors $r = (X^t - X)/(X^t - X^b)$ along minimisation iterations for both pixel and wavelet based distances in presence of an homogeneous observation error.

Comment: Even though $\mathbf{D_A}$ is only an approximation of $\mathbf{R}$, accounting for some part of the spatial correlation is clearly beneficial.

Results 2: Inhomogeneous observation error

The observation error is still Gaussian and spatially correlated, but these correlations are now inhomogeneous in space:

$$Y_{t_i} = Y_{t_i}^t + \epsilon \quad \text{with } \epsilon = \mathbf{A}^T\mathbf{D_A}^{1/2}\beta \quad \beta \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

so that $\mathbf{D_A}$ is the exact representation of $\mathbf{R}$ in the wavelet space.

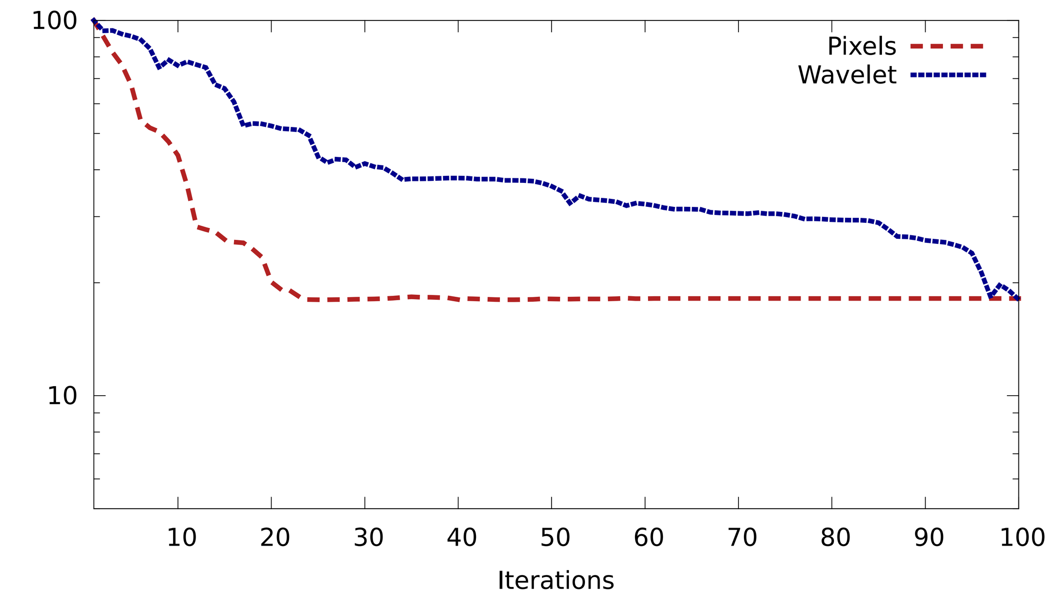

Figure caption: Ratio of residual errors $r = (X^t - X)/(X^t - X^b)$ along minimisation iterations for both pixel and wavelet based distances in presence of an inhomogeneous observation error.

Comment: Even though the observation error is exactly represented in the Wavelet case, its minimisation is struggling to converge toward a minimum similar or better than for the pixel case. Main reason for this: some sort of aliasing in the small scales, as illustrated below:

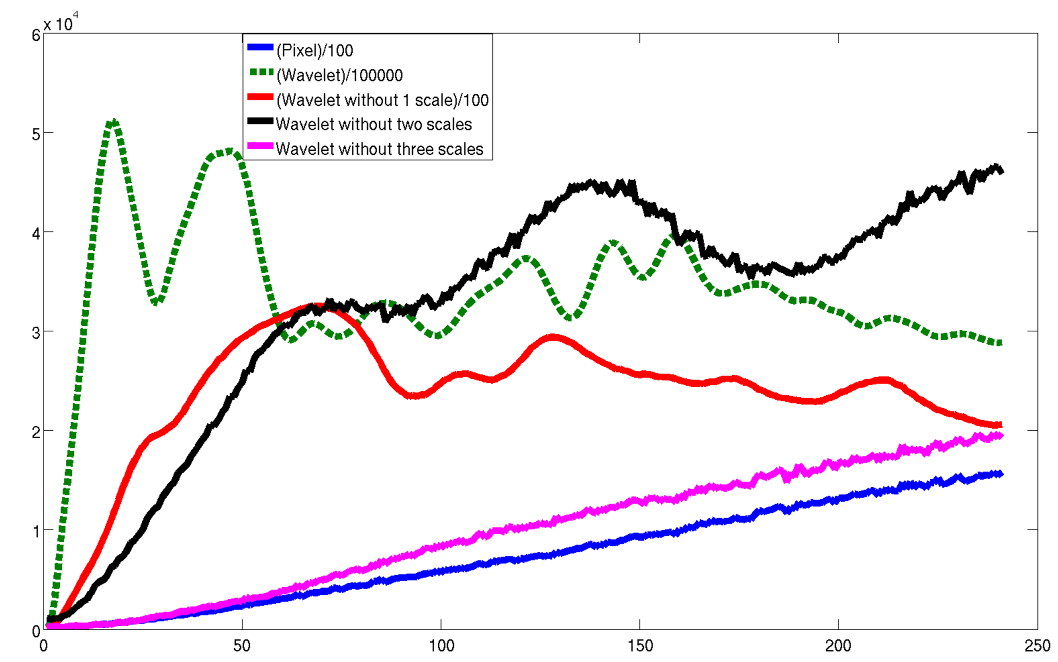

Figure caption: Discrepancy between the background initial concentration and the successive observations $\|Y_{t_i} - \mathcal{H}(X_0^b)\|_{\mathbf{X}}^2$ along time, as would be measured by the observation term of the cost function, with $\mathbf{X} = \text{diag}(\mathbf{R})$ for Pixel and $\mathbf{X} = \mathbf{ARA}^T$ for Wavelet.

Comments:

- Blue / pixel: It starts with a small value (the only difference comes from the noise) and, as time goes by, the vortex drifts and the difference with the initial concentration steadily increases.
- Green / wavelet: It shows a steep increase at the beginning, but then oscillate around a 'plateau'. This happens because, at this point, the norm is really dominated by the small scales. This is expected, since they are the least affected by the correlated noise, so their associated error variances are the smallest (i.e. one trusts more the small scales).
- Red, black and purple / wavelets + removing the 1, 2 and 3 finest scales in the multi scale decomposition respectively. The problem appears later (i.e. for larger discrepancies) when removing the finest scales and even disappear for the purple one.

Results 3: Improvement by online scale selection

Idea: First assimilate the large scales, in order to reduce the difference between $\mathcal{H}(X)$ and Y, and to be in the monotonic region of the green curve, and then progressively includes the smaller scales.

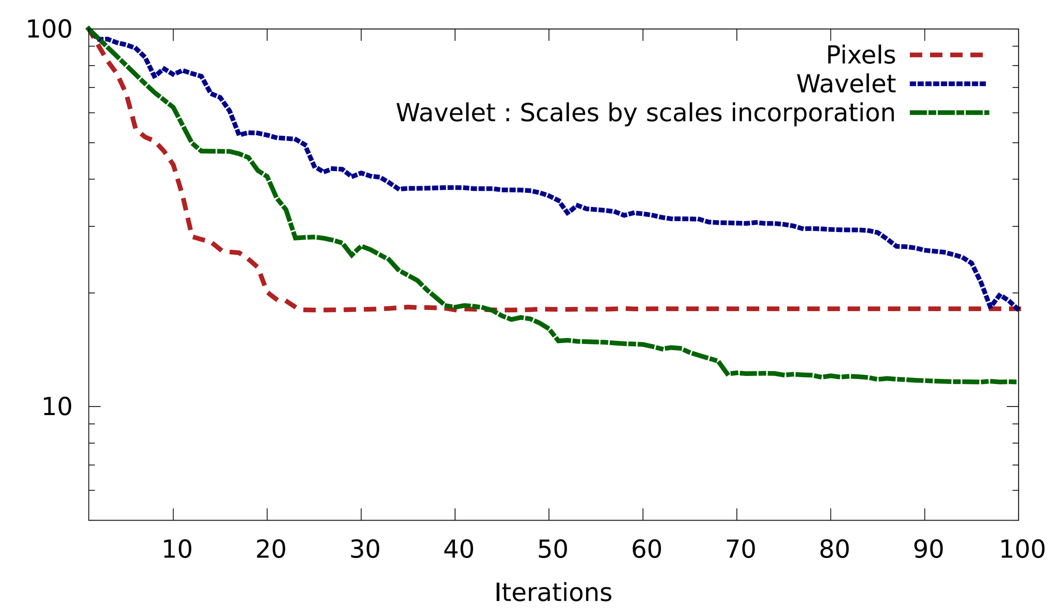

Figure caption: Ratio of residual errors along minimisation iterations for pixel and wavelet-based and progressive wavelet-based distances in presence of an inhomogeneous observation error.

Conclusion: Using multi-scale transforms as changes of variables to represent observation error correlation in data assimilation is a cost effective and promising approach. However this can lead to convergence problems in some cases. A progressive assimilation of the finest scale can significantly improve the convergence, making it a more robust approach.

References

- V. Chabot, M. Nodet, N. Papadakis, and A. Vidard, Accounting for observation errors in image data assimilation, Tellus A, vol. 67, no. 0, pp. 4117?19, Feb. 2015.
- V. Chabot, M. Nodet, and A. Vidard, Progressive assimilation of multiscale observations, IEEE Transactions, 2016.

Papers online: https://team.inria.fr/airsea/maelle/#publi

Acknowledgments: ANR project Geo-FLUIDS