Progressive assimilation of multiscale observations

Vincent Chabot, Arthur Vidard, Maëlle Nodet

Progressive assimilation of multiscale observations

Vincent Chabot
Météo-France
Email: Vincent.Chabot@meteo.fr

Maëlle Nodet
Univ. Grenoble Alpes, CNRS, Inria, LJK
Grenoble F-38000, France
Email: Maelle.Nodet@inria.fr

Arthur Vidard
Inria, Univ. Grenoble Alpes, CNRS, LJK
Grenoble F-38000, France
Email: Arthur.Vidard@inria.fr

***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. In this paper we propose an online scale selection algorithm that improves the convergence properties in such case.**

I. Introduction

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. In this paper, after a short introduction to the context (section II and III), we present, through a simple case mimicking a laboratory experiment (presented in section IV), an illustration of the above-mentioned problem and a possible solution using scale selection during the assimilation process (section V).

II. General formulation of variational data assimilation

Let $\mathcal{M}$ be a dynamical model describing the evolution of the state variable X in space and time:
$$\left\| \begin{array}{l} \partial_t X(X_0, \mathbf{x}, t) + \mathcal{M}(X(X_0, \mathbf{x}, t)) = 0 \\ X(X_0, \mathbf{x}, t_0) = X_0 \end{array} \right.$$
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 $X(X_0^a, \mathbf{x}, t)$ is close to the observations $Y(t)$. This is done by defining X_0^a as the minimum of the cost function:

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

where $\mathcal{V}$ is the model state space, $\mathcal{O}$ the observation space and $\mathcal{H} : \mathcal{V} \mapsto \mathcal{O}$ 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 $\|.\|^2_{\mathcal{V}} = \|.\|^2_{\mathbf{B}}$ and $\|\|^2_{\mathcal{O}} = \|.\|^2_{\mathbf{R}}$ with $\left\|X\right\|^2_{\mathbf{K}} = X^T\mathbf{K}^{-1}X$ where $\mathbf{R}$ and $\mathbf{B}$ are the observation and background error covariance matrices respectively.

III. Error covariance matrix representation

The choice of $\mathbf{B}$ and $\mathbf{R}$ is crucial in data assimilation, since it will drive the way information is spread and how redundancy of error is delt with. So far a strong research effort has been targeted to the background error statistics representation either using multi scale decomposition ([1], and references therein), recursive filters ([2]) or general diffusion models ([3]). In this paper, we mostly focus on the observation error statistics representation, while $\mathbf{B}$ is constructed using the general diffusion approach.

The $\mathbf{R}$ matrix, however, has mostly been assumed diagonal. The main reason is that it simplifies greatly its management. Indeed the number of observations is in general non constant over time which prevents from using the same $\mathbf{R}$ matrix at each assimilation cycle. As a consequence, at each assimilation cycle, a new $\mathbf{R}$ matrix should be formed and inverted. Even without considering the formation of the $\mathbf{R}$ matrix, it size makes its storage and its inversion very difficult.

In [4] we proposed to use linear changes of variable $\mathbf{A}$ to allow for a simpler representation of the observation error covariances. Namely, assuming observation errors to be additive, unbiased and gaussian, i.e. $Y = Y^t + \epsilon$ with $\epsilon \sim \mathcal{N}(0, \mathbf{R})$, Y^t being the true signal. Then $\mathbf{A}Y = \mathbf{A}Y^t + \beta$ with $\beta \sim \mathcal{N}(0, \mathbf{ARA}^T)$.

By choosing $\mathbf{A}$ such that $\mathbf{D_A} = \mathrm{diag}(\mathbf{ARA}^T) \simeq \mathbf{ARA}^T$. one can retain a diagonal approximation, indeed after a bit of algebra, one gets

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

The existence of such $\mathbf{A}$ is no guaranteed, depending on the observation nature and its associated errors. For dense gridded observation with spatially correlated Gaussian errors, [4] used multi-scale transforms, such as Fourier, curvelet and wavelet, for designing $\mathbf{A}$. Doing so, even with a diagonal covariance matrix, one can consider more, or less uncertainties for a given scale compare to the others, while without this change of variable one can only affect grid-point uncertainties. In this paper we will use orthonormal wavelet transform for the operator $\mathbf{A}$ (see [5]). The next sections present two cases study, one where such approach shows clear improvement compare to the classical one, and a second one where it renders the minimisation hieratic. In that case an algorithm based on online scale selection is proposed to circumvent this problem.

IV. Experimental settings

The experimental framework mimics the drift of a vortex on a turntable. The evolution of a vortex in the atmosphere is simulated at the CORIOLIS experimental turntable (Grenoble, France) which re-creates the effect of the Coriolis force on a thin layer of water. A complete rotation of the tank takes 60 seconds which corresponds to one Earth rotation. The vortex is created by stirring the water and made visible thanks to the addition of a passive tracer (fluorescein). The camera is placed above the turntable, and photographs of the vortex constitute the observed image sequence. For more details about these experiments, see [6].

A. Numerical configuration

In this configuration, the evolution of the fluid can be represented by the shallow-water equations involving the horizontal velocity $\boldsymbol{w}(\mathbf{x},t) = (u(\mathbf{x},t), v(\mathbf{x},t))$, where u and v are the zonal and meridional components of the velocity, and the water elevation $h(\mathbf{x},t)$. These unknown variables are defined on the spatial domain $\Omega \ni \mathbf{x}$ and the time interval $[t_0, t_f] \ni t$. Such a model reads:

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

The relative vorticity is denoted by $\zeta = \partial_x v - \partial_y u$ and the Bernouilli potential by $B = g^* h + \dfrac{u^2+v^2}{2}$, where g^* is the reduced gravity. The Coriolis parameter on the β-plane is given by $f = f_0 + \beta y$, κ is the diffusion coefficient and r the bottom friction coefficient. The following numerical values were used for the experiments: $r = 9.10^{-7}$, $\kappa = 0$, $f_0 = 0.25$, $g = 9.81$ and $\beta = 0.0406$.

B. Observation operator

The vortex temporal evolution is shown through the fluorescein concentration evolution. This evolution is observed by an image sequence of the concentration of a passive tracer q transported by the velocity field.

Denoting $\boldsymbol{w}$ the velocity (computed by the model $\mathcal{M}$) transporting the passive tracer and ν_T the diffusion coefficient, we have

$$\left\{ \begin{array}{l} \partial_t q + \nabla q \cdot \boldsymbol{w} - \nu_T \Delta q = 0 \\ q(t_0) = q_0. \end{array} \right. \quad (1)$$

Assuming that the initial concentration of q is known at time t_0, the dynamic of q on the time interval $[t_0; t_f]$ is defined by the model (1), where the diffusion coefficient is $\nu_T = 10^{-5}$.

In the following experiments the considered observation sequence Y represents full maps of q. As a consequence, the observation operator reads :

$$\mathcal{H}\left(X_{t_i}\right) = q(t_i). \quad (2)$$

where $q(t_i)$ comes from (1).

In those experiments we assume that the initial concentration of the passive tracer is known. Therefor we do not control q_0.

C. Twin Experiments Configuration

In order to focus on the methodological aspects we will use a so-called twin experiment framework. In this classical approach, synthetic observations are 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. The result of this analysis can be compared with the synthetic truth. Unless otherwise stated, the assimilation period will be of 144mn, with one snapshot of passive tracer concentration every 6mn (24 snapshot in total). A selection of such observations is shown in fig 1.

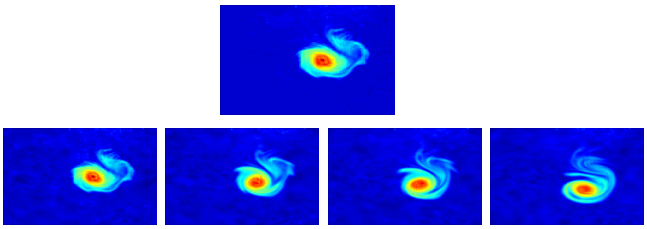

Fig. 1. "True" initial concentration of the passive tracer (top) and noisy observations at initial time, after 90mn, 150mn and 270mn (bottom)

V. Experimental results

A. Effect of the nature of the errors

Two experiments are presented here, with two different kind of observation errors. The first one, with homogeneous Gaussian errors mimic what was done in [4]. The second one introduces more complex inhomogeneous observation errors. In both experiments, two minimization will be performed: a) the comparison between observation and its model equivalent is done grid point by grid point (hereafter Pixel) b) the same comparison is done wavelet coefficient per wavelet coefficient (hereafter Wavelet).

1) Homogeneous observation error: The observation are obtained by adding a spatially correlated Gaussian noise to selected snapshot of the true trajectory

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

In order to mimic the usual approach, only diagonal approximations of $\mathbf{R}$ are used: $\mathbf{R}_{pix} = \mathrm{diag}(\mathbf{R})$ for the Pixel minimisation and $\mathbf{R}_{wav} = \mathbf{D_A} = \mathrm{diag}(\mathbf{ARA}^T)$ Figure 2

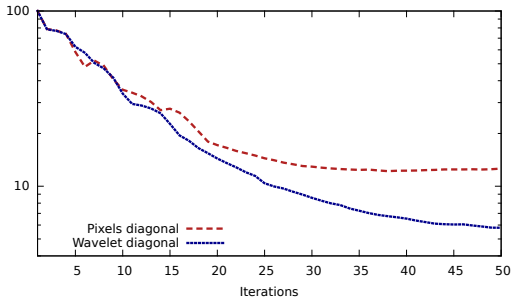

Fig. 2. Ratio of residual errors along minimisation iterations for both pixel and wavelet based distances in presence of an homogeneous observation error

shows the residual error $r = \dfrac{X^t - X}{X^t - X^b}$ along the iterations of minimisation for both Pixel and Wavelet. It shows that, event though $\mathbf{D_A}$ is only an approximation of $\mathbf{R}$, accounting for some part of the spatial correlation is clearly beneficial

2) Inhomogeneous observation error: In the second experiment, the observation error is still Gaussian and spatially correlated, but these correlations are now inhomogeneous in space. For the sake of clarity, this inhomogeneous noise is actually generated out of $\mathbf{D_A}$, through:

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

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

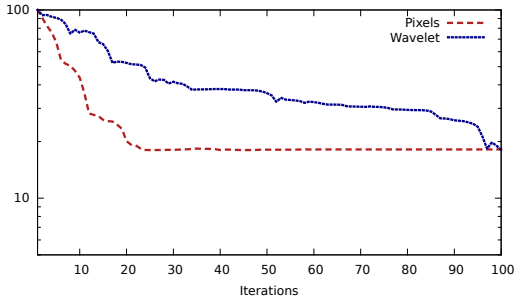

Fig. 3. Ratio of residual errors along minimisation iterations for both pixel and wavelet based distances in presence of an inhomogeneous observation error

Figure 3 shows the same residual error as for figure 2 for a spatially inhomogeneous observation error. 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. The conditioning of the minimisation is probably affected, but the main reason comes from some sort of aliasing in the small scales, as illustrated by figure 4. It represents the norm of the difference between the background concentration and the subsequent observations $\|Y_{t_i} - \mathcal{H}(X_0^b)\|^2_\mathbf{X}$ for $0 \leq i \leq 240$, with $\mathbf{X} = \mathrm{diag}(\mathbf{R})$ for Pixel and $\mathbf{X} = \mathbf{ARA}^T$ for Wavelet. On the one hand, the blue line represents this norm for the Pixel case. 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. As one would expect the farther the vortex drift, the higher the difference with the initial concentration is, all the scales being given the same uncertainties. On the other hand, the wavelet-based norm (in green), 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) and it is the inverse of the variances that is used as a weight in the norm. However it prevents to discriminate between two large scale signals, when the difference is too large (when the green curve stop being monotonic), so the minimisation problem becomes ill-posed. Red, black and purple curves show the same quantity as the green one, but 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. Note that this problem arose as well for the homogeneous case, but after a longer time period (beyond the assimilation window) and in a less striking manner: the 'plateau' is less oscillating and still slightly monotonic.

A solution would be to 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. The next section presents a way to progressively assimilate the smaller scales in order to circumvent the convergence problem.

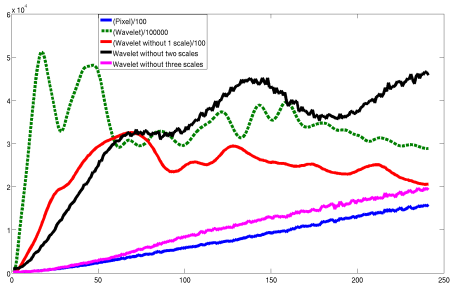

Fig. 4. Discrepancy between the background initial concentration and the successive observations along time, as would be measured by the observation term of the cost function

B. Online scale selection

A simple way to sort out the above mentioned problem would be to remove the smaller scales altogether, but it would mean getting rid of an important part of the information. Instead we propose to gradually include the relevant scales. The contribution J_s to the total cost from a given observed scale can be written

$$J_s(X) = \frac{1}{2n_s} \sum_k \frac{(d^Y_{s,k} - d^{\mathcal{H}(X)}_{s,k})^2}{\sigma^2_{s,k}}$$

where $d_{s,k}$ is a value of wavelet coefficient at scale s and $\sigma^2_{s,k}$ the associated variance (i.e. the corresponding diagonal element in $\mathbf{D_A}$. One can consider that the information at a given scale is usable only if $J_s \leq \tau_s$ so the modified cost

function would replace J_s, from each scale s by

$$J_{s,\tau_s}(X) = \begin{cases} J_s(X) & \text{if } J_s(X) \leq \tau_s \\ \tau_s & \text{otherwise} \end{cases}$$

The green curve in figure 5 shows the evolution of residual error for such modification, with $\tau_s = 4.5$ for all s. This value has been chosen to retain Gaussianity in the retained scales. Indeed, for a Gaussian signal 99% of the considered population should lie within 3 std dev of the mean (here it is a square and divided by two, hence 4.5).

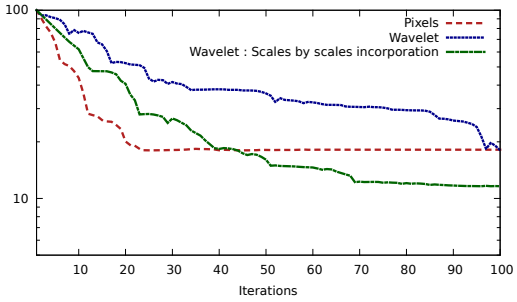

Fig. 5. Ratio of residual errors along minimisation iterations for pixel and wavelet-based and progressive wavelet-based distances in presence of an inhomogeneous observation error.

Figure 6 shows the number of observed snapshot for which each scale is activated (i.e. $J_{s,\tau_s} \neq \tau_s$). All the scales but the finest are gradually included for all snapshots after about 20 iterations. The finest scale is only activated for half of the snapshots after 200 iterations, showing that this information is very difficult to account for by the system.

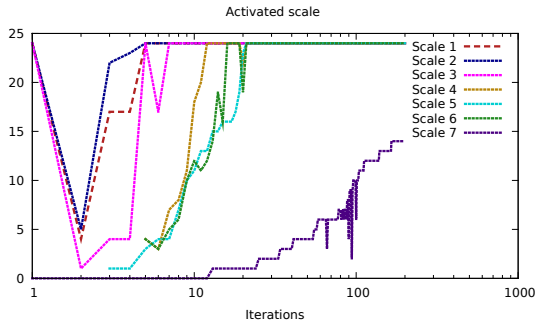

Fig. 6. Number of snapshot for which a given scale is activated along iterations. Scale 1 being the soarse approximation and scale 7 the finest scale.

Figure 7 shows the contribution to the observation term of the cost function from each activated scale. The coarser scales are dominating at the very beginning of the minimisation and converge quite quickly (after 10 iterations), then scales 5 and 6 dominates and converge after 100 iterations. Scale 7 is still including snapshot and has not converged yet after 200 iterations.

VI. 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.

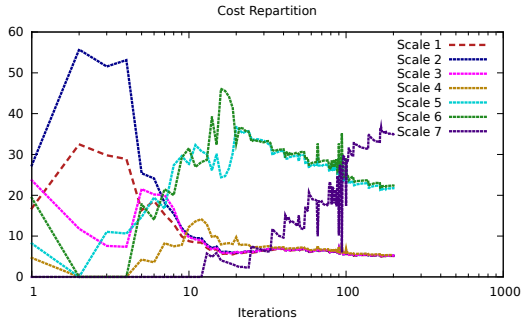

Fig. 7. Contribution to the observation term in the cost function represented by each activated scale along the minimization iteration

Acknowledgment

This work got support from the French national research agency (ANR) through contract ANR-09-COSI-005 (GeoFluids).

References

[1] L. Berre and G. Desroziers, "Filtering of Background Error Variances and Correlations by Local Spatial Averaging: A Review," *Monthly Weather Review*, vol. 138, no. 10, pp. 3693–3720, Oct. 2010.

[2] J. R. Purser, "The filtering of meteorological fields," *Journal of climate and applied meteorology*, vol. 26, no. 12, pp. 1764–1769, 1987.

[3] A. T. Weaver and P. Courtier, "Correlation modelling on the sphere using a generalized diffusion equation," *Q.J.R. Meteorol. Soc.*, vol. 127, no. 575, pp. 1815–1846, Jul. 2001.

[4] 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.

[5] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way* . Academic Press, Oct. 2010.

[6] J.-B. Flór and I. Eames, "The dynamics of monopolar vortices on a topographic beta-plane." *Journal of Fluid Mechanics*, vol. 456, pp. 353–376, 2002.