Accounting for correlated observation errors in image data assimilation

Vincent Chabot, Arthur Vidard, Maëlle Nodet, Nicolas Papadakis

Accounting for correlated observation errors in image data assimilation

Vincent Chabot[*,1,2] — Arthur Vidard[*,1] — Maëlle Nodet[*,2] — Nicolas Papadakis[**]

[*] (1)INRIA and (2)Grenoble Universités,
Laboratoire Jean Kuntzmann 51 rue des Mathématiques, 38041 Grenoble cedex 09
vincent.chabot@imag.fr,arthur.vidard@imag.fr,maelle.nodet@inria.fr

[**] CNRS
Institut de Mathématiques de Bordeaux
351 Cours de la Libération, 33405 TALENCE Cedex, France
nicolas.papadakis@math.u-bordeaux1.fr

ABSTRACT. Satellites images can provide a lot of information on the earth system evolution. Although those sequences are frequently used, the importance of spatial error correlation are rarely taken into account in practice. This results in discarding a huge part of the information content of satellite image sequences. In this paper, we investigate a method based on wavelet or curvelet transforms to represent (at an affordable cost) some of the observation error correlation in a data assimilation context. We address the topic of monitoring the initial state of a system through the variational assimilation of images corrupted by a spatially correlated noise. The feasibility and the reliability of the approach is demonstrated in an academic context.

RÉSUMÉ. Les images satellites sont une source importante d'information sur l'évolution du système terre. Bien que ces séquences d'images soient de plus en plus utilisées, l'importance des corrélations spatiales entre les erreurs présentes en leur sein est rarement prise en compte en pratique. Cela conduit à une sous utilisation de l'information contenue dans ces données. Dans cet article, une nouvelle manière (peu coûteuse) d'intégrer cette information dans le cadre de l'assimilation de données est proposée. Le problème de l'utilisation d'images corrompues par un bruit fortement corrélé en espace afin de contrôler l'état initial du système est abordé. La faisabilité et la pertinence de l'approche proposée est démontrée dans le cadre d'une configuration académique.

KEYWORDS : Data assimilation, covariance modelling, wavelet, curvelet

MOTS-CLÉS : Assimilation de données, modélisation de covariance, ondelettes, courbelettes

1. Introduction

One of the problems in numerical weather prediction is the determination of the initial state of the system. Indeed, the true state of the atmosphere and ocean, at a given moment and in all points of space, are 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, various satellites were put into orbit in order to increase our knowledge of the atmosphere and ocean. Geostationary satellites produce, among other things, sequences of images showing the dynamical evolution of certain types of weather and ocean objects such as clouds, fronts, vortices, ...

Currently, in numerical models of weather prediction, the information contained in this type of sequences is considered through AMV (Atmospheric Motion Vector). They are pseudo-observation of wind velocities derived from satellite image sequences using cross-correlation techniques. However, the error due to the pre-processing of images usually can not be quantified accurately and the information in the estimated velocity fields is therefore difficult to use in practice.

More recently several researches led to propose ways of assimilating images avoiding the use of pseudo-observation. Among them one can cite [3] where they incorporate optical flow constraint directly in the optimality system and [4] where a sequence of model equivalent to the image is created and compared to the observed image. However the description of observation error statistics in this framework has been overlooked so far. In the present paper, an extension of this method is presented in section 3 using several ways to quantify the difference between model and observation. Then the robustness to correlated observation error is then studied in an academic context in section 4.

2. General formulation of variational data assimilation

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

$$\left\| \begin{array}{l} \partial_t \mathrm{X}(\mathrm{X}_0, \boldsymbol{x}, t) + \mathcal{M}(\mathrm{X}(\mathrm{X}_0, \boldsymbol{x}, t)) = 0 \\ \mathrm{X}(\mathrm{X}_0, \boldsymbol{x}, t_0) = \mathrm{X}_0 \end{array} \right. \tag{1}$$

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 $\mathrm{X}(\mathrm{X}_0^a, \boldsymbol{x}, t)$ is close to the observations $Y(t)$. This is done by defining X_0^a as the minimum of the cost function:

$$
\begin{aligned}
J(\mathrm{X}_0) &= J_b(\mathrm{X}_0) + J_o(\mathrm{X}_0) \qquad (2)\\
&= \frac{1}{2}\|\mathrm{X}_0 - \mathrm{X}_0^b\|^2_{\mathcal{V}} + \frac{1}{2}\int_{t_0}^{t_f} ||Y(t) - \mathcal{H}(\mathrm{X}(\mathrm{X}_0, x, t))||^2_{\mathcal{O}}\, dt,
\end{aligned}
$$

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 as advocated by [2].

Typically in data assimilation one uses the Mahalanobis distance $\|.\|^2_{\mathcal{V}} = \|.\|^2_B$ and $\|\|^2_{\mathcal{O}} = \|.\|^2_R$ with

$$
\left\|\mathbf{X}\right\|^2_R = \mathbf{X}^T R^{-1} \mathbf{X} \qquad (3)
$$

where R and B are the observation and background error covariance matrices respectively.

The choice of B and R is crucial in data assimilation. In this paper B is constructed using the general diffusion approach advocated by [6]. R will depend on the way the observation and the model are compared and this is the actual focus of the paper.

Note 1. *For real application, the size of the R matrix makes its storage and its inversion very difficult. Consequently, in this study we will only focus on diagonal approximations (whatever the considered distance). This important to design tractable methods*

3. Image observation operator and associated metrics

As in [4] we assume that we can create q^{synth} a model equivalent to q^{obs} the observed image (an example of such possibility is given in section 4.2).

To perform data assimilation one requires the definition of a distance between observed images and model output. In this paper we investigated several possibilities:

– a pixel to pixel comparison (L^2 norm):

$$
\begin{aligned}
J_o(\mathrm{X}_0) &= \frac{1}{2}\int_{t_0}^{t_f} ||Y(t) - \mathcal{H}(\mathrm{X}(\mathrm{X}_0, \boldsymbol{x}, t))||^2_{\mathcal{O}}\\
&= \sum_{t_0}^{t_f} \|q^{obs}(t) - q^{synth}(t)\|^2_{R_{Pix}} \qquad (4)
\end{aligned}
$$

– Comparison of the image gradients (H^1 seminorm):

$$
J_o(\mathrm{X}_0) = \sum_{t_0}^{t_f} \left\|\nabla q^{obs}(t) - \nabla q^{synth}(t)\right\|^2_{R_{Grad}} \qquad (5)
$$

with ∇ the spatial gradient operator computed using a second order centred finite difference scheme. This approach is motivated by the fact that the information about the dynamics of the system within the tracers sequence lies mainly in its discontinuities.

– Comparison using multi-scale decompositions.

$$J_o(\mathrm{X}_0) = \sum_{t_0}^{t_f} \|W(q^{synth}(t)) - W(q^{obs}(t))\|^2_{R_W} \tag{6}$$

where W is the chosen multi-scale decomposition operator. In this paper two such decompositions are studied namely a wavelet and a curvelet decomposition. The chosen wavelet decomposition is the classical 8 vanishing moments Daubechies and curvelets are described in [1].

In the following we will refer to equations (4,5,6) as R_{pix}, R_{grad}, R_W and R_C respectively (the last two representing wavelet and curvelet decomposition).

4. Twin experiments results

4.1. Experimental context

In order to assess the robustness of the proposed distances, we performed so-called twin experiments where synthetic observations are created using a model output. Subsequently we apply a correlated noise to these synthetic observation and use them in an assimilation experiment. Doing so we have a perfect knowledge of the 'truth' and of the observation error statistics.

In these experiments the dynamics of the system are described by the shallow water equations:

$$\left\{ \begin{array}{lcl} \partial_t u - (f+\zeta)v + \partial_x B &=& -ru + \kappa\Delta u \\ \partial_t v + (f+\zeta)u + \partial_y B &=& -rv + \kappa\Delta v \\ \partial_t h + \partial_x(hu) + \partial_y(hv) &=& 0. \end{array} \right. \tag{7}$$

where $\zeta = \partial_x v - \partial_y u$ is the relative vorticity, $B = g^* h + \frac{u^2+v^2}{2}$ is the Bernouilli potential, for which g^* is the reduced gravity, f is the Coriolis parameter, κ the diffusion coefficient and r the bottom friction coefficient.

The 'true' initial condition (u_0^t, v_0^t, h_0^t) showed in figure 1 simulates the evolution of an isolated vortex in a laboratory experiment using the CORIOLIS [1] turntable.

1. http://coriolis.legi.grenoble-inp.fr

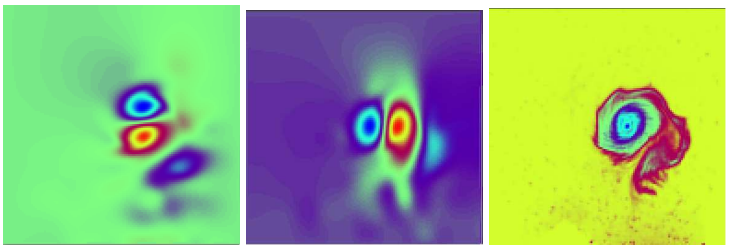

Figure 1. *Left and centre show the zonal (u_0^t) and meridional (v_0^t) components of the 'true' initial velocity field. Right is the first image of passive tracer concentration (q_0^t).*

4.2. Synthetic observations generation

In order to create this synthetic image sequence one assume that the observes quantities correspond to passive tracers that are advected by the dynamic of the system following:

$$\partial_t q + u\partial_x q + v\partial_y q - \nu_T \Delta q = 0 \tag{8}$$

where u and v are the zonal and meridional component of the velocities computed from $\mathcal{M}$, and ν_T is a diffusion coefficient. Equation (8), assuming $q^{synth}(t_0)$ is available, allows to produce an image sequence $q^{synth}(t_i)$ which can in turn be compared to observed images $q^{obs}(t_i)$. Both equations (7) and (8) are tuned to mimic real-life behavior.

However, as mentioned above, in a twin expriment framework the observations also are synthetic. They are generated by transporting the 'true' initial passive tracer concentration q_0^t via equation (8) and using the velocity fields coming from a model simulation starting from the 'true' initial condition. In our academic case, both equations are at the same resolution, but for more realistic applications, one would have to solve (8) at the image resolution (hence interpolating u and v).

Once this sequence of 'true' concentration fields is obtains one add a correlated-in-space Gaussian noise, simulating the effect of observation error. In this paper we used three different level of noise. An example of the effect of these noises is shown in figure 2.

In our data assimilation experiments, only indirect observations (passive tracer concentrations q^t) are used while the model initial condition (u_0, v_0, h_0) is controlled.

4.3. Experimental results

In order to evaluate the effect of the choice of J_o on the quality of the results, the twin experiments described in section 4.1 are performed using observations presented in section 4.2. The starting point (hereafter the background) of the minimisation problem (2) is the system at rest, i.e. $(u_0^b, v_0^b, h_0^b) = (0, 0, h_{mean})$.

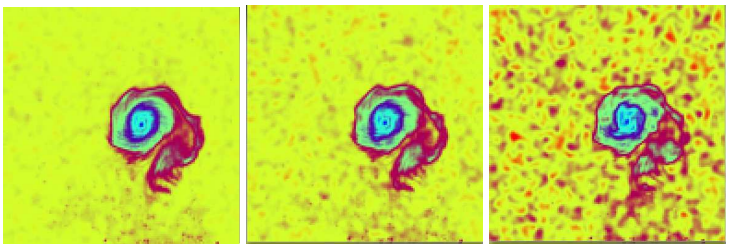

Figure 2. *Examples of noisy observations for each studied noise levels. The corresponding Signal to Noise Ratio (SNR) are from left to right 26.8 dB, 20.8 dB and 14.8 dB respectively.*

For every J_o mentioned in section 3, one needs to described the observation error covariance matrix, bearing in mind that, we will only used diagonal approximations.

R **in pixels' space:** Even though the noise is correlated in space, its repartition within the image sequence is homogeneous in space and time. Consequently each pixel's variance is identical. Therefore if we use in (4) the diagonal of the complete covariance matrix R_{Pix}^{true}, we get

$$R_{Pix} = \text{Diag}(R_{Pix}^{true}) = \sigma^2 \text{I}_n. \tag{9}$$

R **in gradients' space:** In addition to being homogeneous, the used noise presents isotropic correlations. Consequently the diagonal of the complete covariance matrix used in (5) is:

$$R_{Grad} = \text{Diag}(R_{Grad}^{true}) = \tilde{\sigma}^2 \text{I}_{2n}. \tag{10}$$

Note 2. *Let $I_x(i,j)$ the x-direction component of the gradient, its variance is:*

$$\begin{aligned}\tilde{\sigma}^2 = V(I_x(i,j)) =& V\left(\frac{I(i+1,j) - I(i-1,j)}{2}\right) \\ =& \frac{V(I(i+1,j)) + V(I(i-1,j))}{4} - \frac{Cov(I(i+1,j), I(i-1,j))}{2} \\ =& \frac{\sigma^2}{2}(1 - Cor\,(I(i+1,j), I(i-1,j)))\end{aligned}$$

where $Cor(I,J)$ is the correlation coefficient between I and J. This variance depends strongly on the correlations within the observed signal itself. The larger this correlation the smaller $I_x(i,j)$ variance. Therefore a correlated noise would have less impact in the gradient space.

R **in a multi-scale basis:** Knowing the covariance matrix in the pixels' space it is possible to construct its equivalent in the multi-scale basis. It can be done analytically for the wavelets (see for instance [5]) or can be computed empirically for the curvelets.

	Pixel	Grad	Wavelets	Curvelets
14.8 dB	59.8%	33.3%	9.1%	9.7%
20.8 dB	25.6%	17.3%	7.5%	9.2%
26.8 dB	15.1%	11.8%	7.0%	9.7%
Perfect data	7.6%			

Table 1. *Mean over 10 experiments of the residual error on the zonal component of the analysed velocity u_0^a respect to that of the background u_0^b (in percent). Presented for the three levels of noise and for the proposed distances.*

As mentioned before, in (6) we use:

$$R_W = \text{Diag}\left(W R_{Pix}^{true} W^T\right) \tag{11}$$

Note 3. *There is equivalence with formulation* (4) *when*

$$R_W = W R_{Pix} W^T. \tag{12}$$

However, in this case we restrict the matrices to diagonal approximations and:

$$\begin{aligned} R_W &= \textit{Diag}\left(W R_{Pix}^{true} W^T\right) \\ &\neq W \textit{Diag}\left(R_{Pix}^{true}\right) W^T. \end{aligned} \tag{13}$$

Therefore, here (4) *and* (6) *are not identical and we will show that it allows for better representation of error statistics.*

For a spatially correlated noise, the variance of the corresponding multi-scale decomposition coefficients is strongly scale dependent. Consequently the diagonal matrix R_W defined above is not proportional to the identity. This has important implications as illustrated in figure 3, which shows the covariances in pixels' space when using the diagonal approximations in pixel, wavelet and curvelet space compared to the complete covariances.

4.4. Numerical results

Table 1 presents the residual error on the zonal component of the analysed velocity u_0^a (the result of the minimisation) respect to that of the background. Similar results can be seen for the meridional component (not shown). Note that this quantities are not directly observed.

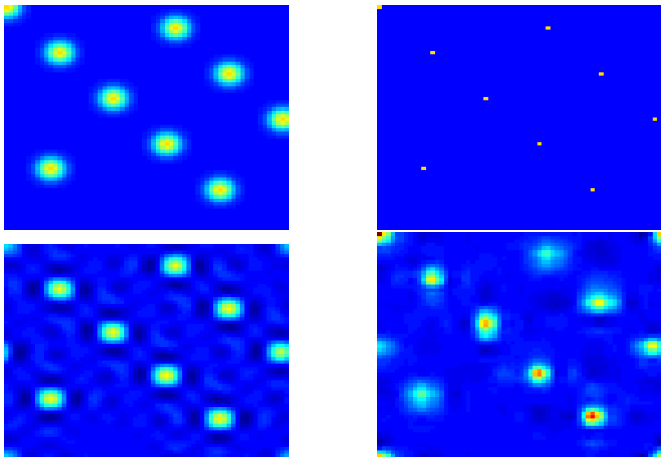

Figure 3. *This figure represents the spatial covariances between 9 individual pixels and their close neighbours in pixels' space. Top left: complete covariances. Top right: using the diagonal approximation* (9) *in the pixels' space. Bottom left: using the diagonal* R_W *in a curvet space. Bottom right: using the diagonal* R_W *in a wavelet space.*

The results obtained for the R_{pix}-norm are shown in the first column. Note that the residual error greatly depends on noise level, which means that the distance between pixels using a diagonal R is not really robust to correlated noise (as expected).

The second column of Table 1 shows the results obtained when using R_{grad}-seminorm. A greater reduction of the error in this case is observed compared to the previous case. These results can be explained if one considers that, "locally", additive correlated-in-space noise introduces a "bias" in the pixels' space. The gradients, by construction are not affected by biases, therefore they are less sensitive to the correlated part of the noise (see note 2).

The performances achieved using a non-proportional to the identity R_W and R_C are shown in the third and fourth columns of table 1. The residual error in both is much lower than in previous two cases. Here the robustness to the level of noise is impressive: the errors in the use of the most noisy sequences are less important than those committed with less noise sequences for other distances. These better results are due to the use of a non-proportional to the identity matrix. Indeed by using different variances for each of the coefficients in the x-let space we are able to represent part of the correlations present in the original signal error.

Fig. 4 shows the analysed fields using the R_{pix}-norm (left) R_{grad}-seminorm (middle) and R_W (right). The velocity field is generally well reconstructed in the vortex area whatever the case (the optimal state being shown in Fig 1 left panel). The main errors for R_{pix} and R_{grad} are in the smooth areas of observation, where the absence of discontinuities

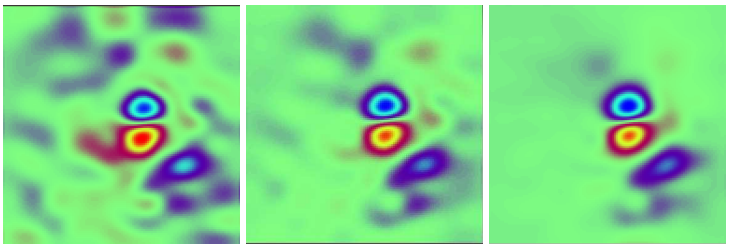

Figure 4. *u component of the analysed velocity field for an observation sequence with strong correlated noise (SNR 14.8 dB). From left to right: using pixels-based distance, using gradient-based semi-distance, using wavelet-based distance.*

in the original image sequence means a lack of information on the underlying movement. The repercussion of the observation error is greater for R_{pix} compared to R_{grad}.

Using multi scale decomposition, it is possible to incorporate some of the noise correlations in the diagonal of the covariance matrix of observation error, this allows to drastically reduce the impact of the noise in smooth areas of the image .

5. Conclusion

In this paper we studied the effect of correlated observation noise on image data assimilation. This was done using four different formulations of the optimisation problem (pixel, gradient, wavelet, curvelet) and applying the usual diagonal approximation of the observation error covariance matrix. We showed that this approximation has a significantly different impact in the different proposed approaches. Indeed, by reconstructing part of the spatial correlation through the multi scale transform, wavelet and curvelet based approaches are almost insensitive to correlated observation noise.

This work was partially funded by the french national research agency through it COSINUS program (Geo-FLUIDS project nº ANR-09-COSI-005).

6. References

Emmanuel Candès, Laurent Demanet, David Donoho, and Lexing Ying. Fast Discrete Curvelet Transforms. *Multiscale Model. Simul.*, 5(3):861, 2006.

Francois-Xavier Le Dimet and Olivier Talagrand. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus A*, 38A(2):97–110, 1986.

Nicolas Papadakis and Étienne Memin. Variational Assimilation of Fluid Motion from Image Sequence. *SIAM J. on Imaging Sciences*, 1(4):343–363, 2008.

Olivier Titaud, Arthur Vidard, I Souopgui, and F.-X. Le Dimet. Assimilation of image sequences in numerical models. *Tellus A*, 62(1):30–47, January 2010.

M. Vannucci and F. Corradi. Covariance structure of wavelet coefficients: theory and models in a bayesian perspective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(4):971–986, 1999.

Anthony T Weaver and Philippe Courtier. Correlation modelling on the sphere using a generalized diffusion equation. *Q.J.R. Meteorol. Soc.*, 127(575):1815–1846, July 2001.