Assimilation of image structures in the presence of observation error

Vincent Chabot, Maëlle Nodet, Nicolas Papadakis, Arthur Vidard

Tellus **000**, 000–000 (0000) Printed 20 December 2013 (Tellus LaTeX style file v2.2)

Assimilation of image structures in the presence of observation error

By V. Chabot[1⋆], M. Nodet[1,3], N. Papadakis[1,2] and A. Vidard[1]

[1] *MOISE team, INRIA Rhônes-Alpes, Campus de Saint Martin d'Hères, 51 rue des Mathématiques, 38041 Grenoble, France.*

[2] *CNRS*

[3] *Université Joseph Fourier, Grenoble, France.*

(Submitted to Tellus 20 December 2013)

ABSTRACT

In this paper, we study the use of image structures as observations in the context of geophysical data assimilation. Instead of assimilating the apparent radiance values of an image sequence to monitor the physical state, we here focus on the shape comparison between a variable of the model and its observation given by images. More precisely, three observation operators based on wavelet transform, image gradient structure and image gradient orientation are proposed and compared with the basic pixel-to-pixel comparison. We also study formally the influence of observation error and its description into the assimilation process in order to define a robust cost function. The effectiveness of the proposed approach is demonstrated on twin experiments performed on a 2D shallow-water model.

Key words: variationnal data assimilation, image assimilation, observation operator, wavelet, approximate covariance matrices, correlated observation errors

⋆ Corresponding author.

e-mail: Vincent.Chabot@imag.fr

1 Introduction

Motivation

In the last thirty years, Data assimilation (DA) techniques have become very popular in geophysics. They aim to combine in an optimal way, through priors on the involved errors, different kinds of information in order to provide an accurate estimate of geophysical state.

From the complexity to describe accurately the dynamic of the ocean or the atmosphere, the mathematical models are in practice idealized and simplified representations of the reality. Observations are then necessary to monitor the evolution of these geophysical states. Observations used in operational systems historically come from synoptic data. Those data are collected by stations, aircrafts, radiosounding, ballons, drifters, ... The repartition of such observations is sparse and heterogeneous both in space and time.

Since the end of the seventies, many satellites have been launched to improve our knowledge of the atmosphere and of the oceans. Geostationary satellites provide, among other data, photographic images of the earth system. Sequences of such images show the dynamical evolution of identified meteorological or oceanic "objects": fronts, clouds, eddies, vortices, etc. The human vision can easily detect the dynamics in this kind of image sequence and it clearly has a strong predictive potential. This aspect is favoured by the fact that this data, contrary to many other measurements, is dense in space and time. Indeed the spatial resolution of the current METEOSAT satellites is close to one kilometre and they produce a full image every 15mn. This frequency will be improved up to one every 10mn (and even every 2.5mn for Europe only) for the upcoming third satellite generation. It implies a huge quantity of information which can be seen as an asset but also induces difficulties for the assimilation system that has to cope with such amount of data.

Satellite data are currently used in operational data assimilation system, mainly through the assimilation of the radiance measured by the satellite at each pixel of the image. They are related to physical quantities such as surface temperature, sea surface height, cloud pressure, chlorophyll concentration... However, study such as Desroziers et al. (2009) have shown that, although radiances measured by satellite have an important impact on assimilation, each radiance observation has a small impact compared to that of one in situ observation. This phenomenon is due to the fact that they are integrated measures. It is thus difficult to have an accurate representation of the observation errors, moreover these errors are highly correlated. As a consequence, the prescribed error statistics associated to radiance measurements are artificially inflated in operational systems.

In practice only a tiny percentage (about $3-5\%$) of total satellite (from polar orbiting and geostationary) data are used in operational Numerical Weather Prediction (NWP) systems and images are only indirectlty used through pseudo observation (see later) and given low confidence within the assimilation systems. Considering the cost of the satellite observing systems (the cost of the launch of the Meteosat Third Generation is estimated at around 2.5 billion Euros) and of the infrastructures required for the collection of the data itself, improving their impact on forecasting systems is an important topic of research.

One important scientific issue related to satellite data is now to study what information can, and should, be assimilated from images. In the literature, several methods dedicated to the assimilation of image structures have been proposed. Two class of approaches can be distinguished. The assimilation of *pseudo–observation* consists in first extracting dynamical information from images and then use the result as observation, whereas the *direct observation* relies on a single assimilation of the image data.

Pseudo–observations

The radiance measured at each pixel is a relevant information, but it does not give any detail on the structures, such as the fronts of geophysical entities, that can be observed in images. When looking at a satellite image sequence, the human eye notices the evolution of structures, through the deformation of isophote lines. Structure information is even available when looking at individual images, as illustrated by the different satellite observations of the ocean presented in Figure 1.

In NWP, this information is currently assimilated through so–called atmospheric motion vectors (AMV). Their aim is to estimate the motion of some identified structures from one image to another using correlation techniques (Lucas and Kanade (1981)). The resulting vector field is then assimilated as wind data. More complex strategies have also been designed by Michel (2011) in order to characterize specific clouds, track their motion along the image sequence and assimilate their Lagrangian trajectories.

Due to their really indirect nature and the complexity of the pre and post processing (a thorough description of these processes can be found in Schmetz et al. (1993) and Nieman et al. (1997)), describing the errors associated to such sparse wind data is not straightforward. In particular they are correlated; so complex observation error covariance matrices have to be built or the errors have to be significantly inflated, and therefore it will reduce their impact. Bormann et al. (2003) found statistically significant spatial error correlations on scales up to about 800 km, that are moreover strongly anisotropic.

Generally AMVs are thought to be very useful. However, as shown by Cardinali (2009), such observations can in some cases have a negative impact on assimilation.

Dense motion estimation have also been considered for assimilating the image dynamics. The motion between two consecutive images can indeed be computed through image processing techniques based on dense optical flow estimation (Horn and Schunck (1981)). The resulting $2D$ vector fields can therefore be associated to the velocity at sea surface or cloud altitude and assimilated as pseudo observation, as proposed by Korotaev et al. (2008) and Papadakis and Mémin (2007).

Other pseudo–observation methods based on image gradients assimilation have been proposed. Gradients contain pertinent information which has been successfully applied for front tracking in oceanographic image sequences (Ba and Fablet (2010)) and assimilation of wave directions (Aouf et al. (2006)).

The pseudo–observation approach nevertheless suffers from the difficulty to model the observation errors. Indeed the error due to the image processing technique itself cannot generally be quantified accurately. For instance, the dense optical flow methods that estimate a 2D velocity from scalar images involve an ill–posed problem and an artificial regularization of the vector field is required. Such regularization introduces errors in the estimation which are difficult to model. As a consequence, when dealing with pseudo–observations, it is hard to distinguish image processing errors from the ones inherent to the satellite acquisition process.

Direct observation

The direct comparison between image temporal variation and the state velocity has been studied by Papadakis and Mémin (2008), through the introduction of a non linear observation equation based on the optical flow constraint equation. This overcomes the problem of artificial regularization by only considering the photometric information. An other approach has also been developed by Titaud et al. (2010) where observed images are compared directly with a synthetic image derived from the model state variable.

More recently, another interesting approach using image gradients has been proposed by Titaud et al. (2011) and continued by Gaultier et al. (2012). The idea here consists in extracting a feature map from the dynamic that can be compared with a gradient map computed on images. The feature map is composed of Lyapunov exponents or vectors (D'Ovidio et al. (2009)) obtained through an integration of the position of particles added to the model. More precisely, the computation consists in perturbing a particle position and measuring the direction and the amplitude of the deviation after a finite integration time. In oceanography, this information is pertinent

since strong Lyapunov coefficients are correlated with chlorophyl discontinuities at the surface, that can be observed from satellite images. The process nevertheless relies on a binarization of both Lyapunov information and image gradients, which makes its integration in data assimilation system difficult, as it involves non differentiable processes and degrades the information.

Specification of observation errors

The specification of observation errors statistics is an important topic in data assimilation. Indeed they define the weight of each observation and the way the analysis can differ from the observation set (through the error covariances).

Estimating the error statistic of an observation set is a mathematically difficult problem and has seldom been considered thoroughly in data assimilation. Desroziers et al. (2005) introduced a method in order to diagnose the consistency of background and observation error variances and cross–correlations under the hypothesis that the assimilation scheme rely on statistical linear estimation theory. On top of this estimation difficulties, introducing the information about cross–correlation in the assimilation scheme is a complex task due to the size of the problem and the need to invert the obtained covariance matrix. To circumvent this complexity, Stewart et al. (2013) proposed to use Markov matrices in order to model correlations. The inverse of such matrices are tri–diagonal and need the storage of only one parameter. They show that, on synthetic test cases, using information on correlation (or even approximation of this information) leads to better results than considering the observation errors uncorrelated.

The specification of observation error for image–type observation is a challenge. Whatever the approach used (pseudo– or direct observation) the usual assumption of uncorrelated errors between observation is clearly invalid. Additionally, the size of the data is likely to make the handling of the error covariance matrices cumbersome. This is one of the main topics of this paper.

Proposed approach and overview of the paper

This paper compares four different observation operators dedicated to image–type observation. Two of them have already been proposed in the literature: one–to–one pixel comparison (Corpetti et al., 2009) and the comparison in a curvelet space (Titaud et al., 2010) extended in this paper to the comparison in a orthonormal wavelet spaces. Two new operators based on image gradient location and orientation are also proposed. Their relative merits are discussed with a special attention on the description of the corresponding error statistics specification.

Some basic notations on Variational Data Assimilation are first recalled in section 2. Observation operators measuring directly the difference between the state variable and the image gradients are then defined in section 3. In section 4 the proposed new approaches are applied to an academic test case with no observation error and compared to the previously proposed operator. The specification of observation error statistics for each operator is discussed in section 5. This is a very important step in order to propose adequate error covariance matrices for assimilation purposes. Finally section 6 discuss the impact of observation noise on the proposed approaches.

2 Variational Data Assimilation

Variational Data Assimilation is a technique based on optimal control theory which seek to estimates initial or boundary conditions of a given system. This is done through the minimisation of a cost function accounting for observation and model equations of this system.

Let $\mathcal{V}$ be the state space identified to its dual defined over $\Omega \subset \mathbb{R}^n$ (e.g. $\mathcal{V} = H^1(\Omega)$). The evolution of the state variable $X \in \mathcal{W}(t_0, t_f) = \{f | f \in L^2(t_0; t_f; \mathcal{V})\}$ is assumed to be described through a (possibly non linear) differential dynamical model $\mathcal{M} : \mathcal{V} \mapsto \mathcal{V}$:

$$\begin{cases} \partial_t X(X_0, t) + \mathcal{M}(X(X_0, t)) = 0 \\ X(X_0, t_0) = X_0 \end{cases} \tag{1}$$

where $X_0 \in \Omega$ is a control parameter. We then consider noisy observations $Y \in \mathcal{O}$ of the state variables, where $\mathcal{O} \subset \mathbb{R}^m$ is the observation space. These observations may belong to a different space from the state variable. We will nevertheless assume that there exists a differentiable operator $\mathcal{H} : \mathcal{V} \mapsto \mathcal{O}$, that goes from the variable space to the observation space. The control problem states as follows: find the control parameter X_0 which fits best to observations over the time range $[t_0; t_f]$. To do so we form and minimize the following cost function:

$$J^o(X_0) = \frac{1}{2} \int_{t_0}^{t_f} ||Y(t) - \mathcal{H}(X(X_0, \boldsymbol{x}, t))||_R^2 \, dt, \tag{2}$$

where R is the observation error covariance matrix and the norm $\|.\|_R$ is defined by

$$\|u\|_R^2 = u^T R^{-1} u \tag{3}$$

for all $u \in \mathcal{O}$. It is classically known that this problem is ill–posed. Therefore a regularization term, involving a so–called background state, is added to the cost function:

$$J(X_0) = J^o(X_0) + J^b(X_0), \quad J^b(X_0) = \|X_0 - X^b\|_B^2 \tag{4}$$

where B is the background error covariance matrix. In this paper it is based on the generalized diffusion equation

as proposed by Weaver and Courtier (2001).

The minimum of J is sought using a gradient descent algorithm and its gradient is generally computed using an adjoint method as advocated by Le Dimet and Talagrand (1986).

3 Assimilation of image structures

The general framework needed to directly assimilate images of flow–driven passive tracer is described in section 3.1. Section 3.2.1 and 3.2.2 recall the pixel observation operator (introduced by Corpetti et al. (2009)) and the multiscale observation operator (proposed by Titaud et al. (2010)). Then sections 3.3 et 3.4 introduce new operators based on image gradients.

3.1 General framework

Let I be an image sequence associated to a fluid flow whose dynamics is driven by the mathematical model $\mathcal{M}$. The observation I can be a direct or indirect measurement of a specific state variable of the mathematical model. For indirect measurements, in order to define a distance between model output and images, Titaud et al. (2010) proposed to create a synthetic image from the model state. The evolution of a tracer q is used in order to synthesize a model equivalent to the image sequence. This tracer is assumed to represent the phenomenon observed in the images I, so that a direct comparison between tracer values and image intensity is possible. In the experimental sections 4 and 6, we assimilate a monochromatic passive tracer. Therefore the link between the passive quantity and the state variables naturally appears from the advection–diffusion equation that defines the tracer dynamics. Denoting $\boldsymbol{w}$ the velocity (computed by the model $\mathcal{M}$) transporting the passive tracer and ν_T the diffusion coefficient, we have

$$\begin{cases} \partial_t q + \nabla q \cdot \boldsymbol{w} - \nu_T \Delta q = 0 \\ q(t_0) = q_0. \end{cases} \tag{5}$$

The initial condition q_0 of the passive tracer is assumed to be known (for example given by the first available image). The corresponding functional to minimize then reads:

$$J^o(X_0) = \int_\Omega \int_{t_0}^{t_f} ||q - I||^2_{\mathcal{F}} \, d\boldsymbol{x} \, dt, \tag{6}$$

where X_0 denotes the unknown state initial conditions that we want to estimate, including for instance the initial state velocity $\boldsymbol{w}_0$. $\mathcal{F}$ represents the space in which the comparison between tracer output and images is performed.

3.2 Structure observation operators

In the following, we will denote as $I(\boldsymbol{x}, t)$ the observed image and $q(\boldsymbol{x}, t)$ the corresponding state variable at the same time. For the sake of simplicity, the variable q and the observation I are assumed to be defined on the same discrete spatial domain Ω composed of $L \times H$ pixels, so that $n = m = |\Omega| = LH$. We now describe several norms $\mathcal{F}$ to compare image and tracer.

3.2.1 Pixel observation operator

A first idea for comparing images consists in defining $\mathcal{F}$ as the pixel space, which will be called *pixel observation operators*. The functional to minimize is:

$$J^o_{pix}(X_0) = \int_\Omega \int_{t_0}^{t_f} ||q - I||^2_R \, d\boldsymbol{x} \, dt, \tag{7}$$

where R stands for the covariance matrix in this space. In fact, this pixel observation operator acts like *in situ* measurements, as it directly measures the value of a state variable.

3.2.2 Multiscale observation operator

In order to model the spatial information, an idea is to consider a multiscale representation of an image. Titaud et al. (2010) proposed to use the curvelet representation defined in Candes et al. (2006) to extract significant structures and filaments contained in the images. The very same process can be done using other multiscale representation. In this paper, we propose to use orthonormal periodic wavelet transform. This choice is due to the fact that wavelets basis involve faster transforms and are easier to work with. The wavelet and curvelet transforms used in this paper comes from MCAlab toolbox[1] (see Fadili et al. (2010)). These transforms rely on the fact that an image I can be decomposed as :

$$I = \sum_{j,\mathbf{k},l} < I, \varphi_{j,\mathbf{k},l} > \varphi_{j,\mathbf{k},l}$$

where $\{\varphi_{j,\mathbf{k},l}\}_{j,\mathbf{k},l}$ are the elements of the curvelet tight frame or of the orthonormal wavelet basis. The parameters $j, \mathbf{k}, l$, specify the scale, the position and the orienation of each element. In order to exhibit the structures from the observed/synthetic images and deal with data noise, a thresholding method is applied on the coefficients of the observed image. These coefficients are then compared with their synthetic counter–part. Hence, the functional to minimize reads:

$$J^o_W(X_0) = \int_\Omega \int_{t_0}^{t_f} ||\mathcal{T}_I \circ W(q) - \mathcal{T}_I \circ W(I)||^2 \, d\boldsymbol{x} \, dt, \tag{8}$$

[1] https://fadili.users.greyc.fr/demos/WaveRestore/downloads/mcalab/Home.html

where W stands for the multiscale transform and $\mathcal{T}_I$ is a thresholding method associated to a threshold $\tau > 0$ (based on the observation coefficients) reading

$$\mathcal{T}_I(\langle q, \varphi_{j,\mathbf{k},l}\rangle) = \begin{cases} \langle q, \varphi_{j,\mathbf{k},l}\rangle \ \text{ if } |\langle I, \varphi_{j,\mathbf{k},l}\rangle| \geqslant \tau \\ 0 \text{ if } |\langle I, \varphi_{j,\mathbf{k},l}\rangle| < \tau \end{cases} \tag{9}$$

To summarize, the comparison is done in a subspace defined through a thresholding operation on the observed data.

Notice that we have a Parseval equality for the wavelet and curvelet transform, so that without thresholding the norm $\|W(q) - W(I)\|_{L^2}$ is equal to the pixel–space $\|q - I\|_{L^2}$ norm.

3.3 Gradient observation operator

Considering that the pertinent information on the flow dynamics is localized on image singularities, we also propose to compare 2D gradients $\nabla = [\partial_x \quad \partial_y]^T$ of both quantities. Hence, the observation operator is defined as $\mathcal{H}_\nabla(q) = \nabla q$. As a result the cost function becomes:

$$J^o_\nabla(X_0) = \frac{1}{2}\int_{t_0}^{t_f}\int_\Omega \|\nabla I - \nabla q\|^2_R \, d\boldsymbol{x}\, dt, \tag{10}$$

where R is the covariance matrix relative to the observation errors $\mathcal{H}_\nabla$. The corresponding tangent linear operator can be obtained for a small perturbation γ around q as:

$$\partial_q \mathcal{H}_\nabla \gamma = \nabla\gamma.$$

The adjoint observation operators applied to a vector u reads:

$$(\partial_q \mathcal{H}_\nabla)^* \rho = -\nabla.\rho$$

Remark 1. There exist an infinity of different tracers that share the same gradient map. Indeed, the family $q^c = q + c$, where c is a global constant, is such that $\nabla q^c = \nabla q$. By fixing q_0 the initial tracer values, we nevertheless ensure the uniqueness of the solution, since the mass of the tracer is preserved along time through the conservation equation (5) (with Neumann homogeneous or periodic boundary conditions).

3.4 Angular observation operator

A last operator can be introduced to focus on structure orientations. The idea is to compare the normalized 2D gradients $\nabla = [\partial_x \quad \partial_y]^T$ of both quantities, by measuring their angular difference. We then measure the misfit as

$$\frac{\nabla q}{||\nabla q||} - \frac{\nabla I}{||\nabla I||},$$

with the L^2 norm: $||\nabla q|| = \sqrt{q_x^2 + q_y^2}$. We can indeed consider that the contour lines of the quantity q are conserved in the image I, even if the acquisition process is non–linear. This happens in practice with satellite acquisitions, where the measurements correspond to a vertical integration of the observed quantity. Such non–linear observations thus lead to very difficult inverse problems. By considering the contour lines of both quantities, we expect to circumvent this limitation. In order to avoid an ill–posed problem for null vectors, we rather define the misfit as:

$$\frac{\nabla q}{||\nabla q||_\epsilon} - \frac{\nabla I}{||\nabla I||_\epsilon},$$

where $||\nabla u||_\epsilon = \sqrt{u_x^2 + u_y^2 + \epsilon^2}$ and $\epsilon > 0$ is chosen as $\epsilon = ||\nabla I||^+/10$, with:

$$||\nabla I||^+ = \min_{\substack{\boldsymbol{x} \in \Omega \\ t \in [t_0, t_f]}} \Big\{ ||\nabla I(\boldsymbol{x}, t)||, \text{ such that } ||\nabla I(\boldsymbol{x}, t)|| > 0 \Big\}. \tag{11}$$

As a consequence, this leads to the following inequality when ∇I is not null:

$$\frac{0.995}{\|\nabla I\|} \leqslant \frac{1}{\|\nabla I\|_\epsilon} \leqslant \frac{1}{\|\nabla I\|}$$

Therefore $\frac{\nabla I}{||\nabla I||_\epsilon}$ is either null or almost equal to the unitary vector $\frac{\nabla I}{||\nabla I||}$. Following the notations of section 2, the observation operator is now defined as $\mathcal{H}_{\text{Ang}}(u) = \frac{\nabla u}{||\nabla u||_\epsilon}$ and the observations are obtained as $Y = \mathcal{H}_{\text{Ang}}(I)$. The corresponding functional reads:

$$J^o_{\text{Ang}}(X_0) = \frac{1}{2} \int_{t_0}^{t_f} \int_\Omega \left|\left| \frac{\nabla I}{||\nabla I||_\epsilon} - \frac{\nabla q}{||\nabla q||_\epsilon} \right|\right|^2_R d\boldsymbol{x}\, dt, \tag{12}$$

where R is the covariance matrix relative to the observations errors. As the observation operator $\mathcal{H}_{\text{Ang}}$ is non linear, we need to compute the tangent linear operator $\partial_q \mathcal{H}_{\text{Ang}}$. It can be obtained for a small perturbation γ

around q as:

$$\partial_q \mathcal{H}_{\text{Ang}} \gamma = \frac{1}{||\nabla q||^3_\epsilon} \underbrace{\begin{bmatrix} q_y^2 + \epsilon^2 & -q_x q_y \\ -q_x q_y & q_x^2 + \epsilon^2 \end{bmatrix}}_{C(q)} \nabla \gamma. \tag{13}$$

The eigenvalues of the symmetric matrix $C(q)$ are ϵ^2 and $||\nabla q||^2 + \epsilon^2$, respectively associated to the eigenvectors ∇q and $\nabla^\perp q$. As a consequence, if $||\nabla q||$ is not zero, most of the information will be send to $\nabla^\perp q$, the orthogonal direction of the current state gradient direction. On the other hand, this operator will just realize an isotropic diffusion if ∇q is a null vector. The adjoint observation operator applied to a vector u reads:

$$(\partial_q \mathcal{H}_{\text{Ang}})^* u = -\nabla \cdot \left(\frac{1}{||\nabla q||^3_\epsilon} \begin{bmatrix} q_y^2 + \epsilon^2 & -q_x q_y \\ -q_x q_y & q_x^2 + \epsilon^2 \end{bmatrix} u \right). \tag{14}$$

The right handside of the adjoint model is finally obtained by taking $u = R^{-1} \left(\frac{\nabla I}{||\nabla I||_\epsilon} - \frac{\nabla q}{||\nabla q||_\epsilon} \right)$ in the previous expression.

Once again there exist an infinite number of different tracers q sharing the same gradient directions, such as the tracer family $q^{\alpha,c} = \alpha q + c$, $\forall \alpha \in \mathbb{R}$ and $c \in \mathbb{R}$. Having the mass preservation relation (5) and fixing q_0 is here not sufficient to avoid any ambiguity. As a consequence, the angular operator is not able to correct the intensity of the tracer and will only focus on correcting the position of structures.

According to the current values of ∇q^k and ∇I, the minimization of (12) produces the following effect on q^{k+1}:

- minimize the norm $||\nabla q^{k+1}||$ in the areas where ∇I is null,
- diffuse the orientation of the observation orientation ∇I where $||\nabla I|| > 0$ and ∇q^k is null,
- maximize the scalar product $\frac{\nabla q^{k+1}}{||\nabla q^{k+1}||_\epsilon} \cdot \frac{\nabla I}{||\nabla I||_\epsilon}$ where ∇I and ∇q^k are non null.

These properties illustrates how the structure orientation of images is transferred to the model variable.

4 Validation of Gradient-based operators in perfect case

4.1 Experimental framework and dynamical model

The experimental framework of the paper concerns the study of a drift of a vortex on a turntable. The evolution of a vortex in the atmosphere is simulated at the CORIOLIS experimental turntable[2] (Grenoble, France) which

[2] http://coriolis.legi.grenoble-inp.fr

re–create 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 (the fluorescein, denoted as q). The camera is placed above the turntable, and photographs of the vortex constitute the observed image sequence. For more details about thoses experiments, see Flór and Eames (2002).

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

$$\left\{ \begin{array}{rcl} \partial_t u - (f+\zeta)v + \partial_x B &=& -ru + \kappa\Delta u \\ \partial_t v + (f+\zeta)u + \partial_y B &=& -rv + \kappa\Delta v \\ \partial_t h + \partial_x(hu) + \partial_y(hv) &=& 0. \end{array} \right. \tag{15}$$

The relative vorticity is denoted as $\zeta = \partial_x v - \partial_y u$ and the Bernouilli potential as $B = g^* h + \frac{u^2+v^2}{2}$, where g^* is the reduced gravity. The Corolis 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: $r = 9.10^{-7}$, $\kappa = 0$, $f_0 = 0.25$, $g^* = 0.02$ and $\beta = 0.0406$.

The simulation is performed on a rectangular domain $[0, L] \times [0, H]$ representing a sub–domain of the turntable with $L = H = 2.525$m. The domain is discretized on a 128×128 uniform Arakawa C–type square grid. A finite differences scheme is used for space discretization. Time integration is performed using a fourth order Runge–Kutta scheme. Time step of the turnable experiment is set to $0.01s$ which would correspond to $14.4s$ in the atmosphere. The vortex temporal evolution is shown through the fluorescein concentration evolution. This evolution is observed by an image sequence whose grey levels correspond to the concentration of the passive tracer q transported by the velocity field.

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 (5), where the diffusion coefficient is $\nu_T = 10^{-5}$. The considered image sequence I represents observations of q.

4.1.2 Twin experiments configuration In order to study the quality of the different observation operators, we will deal with synthetic examples where the groundtruth is known. This groundtruth is used to generate observations. Considering initial conditions at time t_0 of the state variables (u_0^*, v_0^* and h_0^*) and fluorescein concentration q_0^*, a simulation of the dynamical model is realized from $t_0 = 0s$ to $t_f = 6s$. This defines a coherent scenario, as illustrated in Figure 2.

The simulated values of the concentration $q^*(t_i)$, at various observation times $t_i \in]t_0; t_f[$, are then used as observations. We chose to have an image every $0.25s$, so that we finally get 25 observed images.

In the experiments, a VDA algorithm, allowing the estimation of the state variables (u, v, h), is applied on different data, initializing the background variables $u^b(t_0, \boldsymbol{x})$, $v^b(t_0, \boldsymbol{x})$ with null values and $h^b(t_0, \boldsymbol{x}) = 0.3553m$. The minimization of the cost function is realized thanks to N1QN3 solver (see Gilbert and Lemaréchal (1989)).

4.1.3 Metrics for qualitative analysis We here introduce some tools in order to measure different criteria on the estimated velocity fields. We mainly focus on the estimation errors for the velocities u and v. The vorticity error and the angular error are also looked at, as it gives an excellent criteria to evaluate the quality of the estimated velocities directions.

In the following, we refer to the Root Mean Square Error (RMSE) between the estimated variables u and the true synthetic ones u^* at time t_0:

$$RMSE(u) = \sqrt{\frac{1}{|\Omega|} \sum_{\boldsymbol{x} \in \Omega} (u(t_0, \boldsymbol{x}) - u^*(\boldsymbol{x}))^2}.$$

Simlarly, we define the RMSE for v and the vorticity $\zeta = v_x - u_y$, as well as the angle α defined as:

$$\alpha = \arccos\left(\frac{\langle U_\epsilon, U_\epsilon^* \rangle}{||U_\epsilon||.||U_\epsilon^*||}\right), \tag{16}$$

where $U(\boldsymbol{x}) = [u(t_0, \boldsymbol{x}), v(t_0, \boldsymbol{x}), \epsilon]$, $U^*(\boldsymbol{x}) = [u^*(t_0, \boldsymbol{x}), v^*(t_0, \boldsymbol{x}), \epsilon]$ and $||U_\epsilon||^2 = \langle U_\epsilon, U_\epsilon \rangle$. This corresponds to a modification of the angular error of Barron et al. (1994) as proposed in Souopgui (2010).

In the following, we mostly consider the ratio $RMSE(u^a)/RMSE(u^b)$. This ratio represents the residual error of the estimated zonal velocity u^a with respect to the error of the background (or initial) zonal velocity u^b. A ratio smaller than 1 means that the assimilation has improved the corresponding variable with respect to the background. These RMSE and ratios can be computed on the whole domain or on specific areas (e.g. the red box on figure 2).

Remark 2. In our experiments, as the background velocities are initialized with null values, this ratios reads:

$$\frac{\sqrt{\sum_{\boldsymbol{x}\in\Omega}\left(u^a(t_0,\boldsymbol{x})-u^*(t_0,\boldsymbol{x})\right)^2}}{\sqrt{\sum_{\boldsymbol{x}\in\Omega}\left(u^*(t_0,\boldsymbol{x})\right)^2}} \tag{17}$$

Therefore the square of this number represents the percentage of noise (in terms of energy) present in the analyzed field.

4.2 Validation with Perfect data

We want to compare the behavior of *gradients* (see §3.3) and *angular* (see §3.4) observation opertors to *pixel*, *curvelet* and *wavelet* observation operators (see §3.2) in the case of perfect data (i.e without any noise).

As we started with a static background, we have no confidence in it. Therefore, we introduce a weight w_b in the cost function:

$$J = J^o + w_b\, J^b$$

Choosing w_b small ensures that the J^b term acts only as a regularization term, and not as a strong feedback to the background.

Figure 3 shows the evolution of the RMSE ratio of the zonal velocity u along the iterations for *pixel*, *wavelet*, *gradients* and *angular* observations operators. For each distance notion, this ratio decrease along iterations. The minimisation leads to a coherent analyzed field in each case. The best zonal velocities field is reconstructed with the *angular* observation operator.

RMSE ratios obtained at the end of the minimization processes are presented in table 1. As illustrated by Figure 3, most of the initial velocity error is corrected throughout data assimilation. Indeed, for the velocities components (u and v), the noise energy of the analyzed field represent less than 2.5% of the true field energy.

One can notice that the *angular* observation operator leads to slightly better results than other observation operators for all criteria. This could mean that by considering only the orientation of the gradient the pertinent information about the structure presents in the images has been selected. One can see that *wavelet* observation operator leads to better results than *pixel* . Therefore, removing some information does not perturb the assimilation process with perfect data.

In this perfect case, an analyzed field close to the groundtruth state can be estimated for all the considered observation operator. This validates the proposed gradient and angular operators.

However, as the world is imperfect, we now study the impact of different kind of observation errors on the assimilation process in the next sections.

5 Noise modeling

When dealing with real satellite data, noise can perturb the earth surface or atmospheric signal. This noise is due to radar measurement uncertainties as well as perturbations involved by the vertical integration of the signal (as the astmophere can contain clouds, rain, snow, ...). The induced errors must be characterized in order to perform an optimal analysis.

The measurement norm (3) involved in the observation cost function (2) requires the inverse of the observation error covariance matrix R. As these matrices can be huge (and even non invertible) one needs to make some approximations.

As a consequence, in this part we focus on building approximate covariance matrices easily invertible in different space. We would like to underline that in a wavelet, curvelet, gradient and angular space a diagonal covariance matrix can hold information on error correlation in the pixel space (this fact is discussed with more details in §5.3 for the wavelet case). This fact has been already used in a wavelet basis for instance in Pannekoucke (2009) to model heterogeneous correlation in the background error covariance matrice.

5.1 Notations

In the following, the true error covariance matrices are denoted C^{noise}_{Space} depending on the *noise* type and the considered observation *Space*.

Let us first consider that the available images I^o correspond to the true state I^*, pertubed with an additional time independent error η. The observation equation reads, in the pixel space, $I^o = I^* + \eta$. Assume that the error covariance matrix, C^{noise}_{Pix}, is known. Then the error covariance matrix associated to a linear transform A of the observed image is:

$$C^{noise}_{A} = AC^{noise}_{pix}A^T \tag{18}$$

where A can be a multiscale transform described in §3.2.2 or the gradient operator from §3.3.

As these matrices can be huge (and even non invertible) one needs to approximate them $R^{noise}_{Space,Approx} \approx C^{noise}_{Space}$. The approximations considered are:

- Scalar $R^{*}_{*,Scalar} = \sigma^2 \mathrm{I}_n$,
- Diagonal $R^{*}_{*,Diag} = Diag(C^{*}_{*})$,
- Block diagonal $R^{*}_{*,Block}$,

where I_n is the $n \times n$ identity matrix. For instance, when considering an independent and identically distributed (iid) noise, the matrix $R^{iid}_{\nabla,Block}$ is a block diagonal approximation of the true error covariance matrix C^{iid}_{∇} for the gradient space ∇.

5.2 Independent additive noise

Assume that images are corrupted by a gaussian white noise independent and identically distributed in space and time. Let us describe the error covariance matrices build in this case for the operator introduced in section §3.2.1, §3.2.2, §3.3 and §3.4.

Pixel case: The true error covariance matrice C^{iid}_{Pix} is scalar. As a consequence there is no need to approximate this matrix:

$$R^{iid}_{Pix,Scalar} = C^{iid}_{Pix} = \sigma^2 \mathrm{I}_n, \tag{19}$$

where σ^2 monitors the noise amplitude.

Wavelet case: The error covariance matrix in a wavelet space is:

$$C^{iid}_{W} = \sigma^2 W W^T$$

where W is the wavelet transform defined in §3.2.2. As only orthonormal wavelets transforms are used in our study we have:

$$W^T = W^{-1}.$$

Then, whatever the chosen wavelet basis, provided that it is orthonormal, the error covariance matrix is

$$R^{iid}_{W,Scalar} = C^{iid}_{W} = \sigma^2 \mathrm{I}_n. \tag{20}$$

When a thresholding operation $\mathcal{T}_I$ (see (9))is applied on the wavelet coefficients, the covariance matrix is still modelled by a scalar matrix:

$$R^{iid}_{\mathcal{T}_I(W),Scalar} = \sigma^2 \mathrm{I}_k, \tag{21}$$

where k is the number of coefficients kept by the thresholding.

Remark 3. Keeping all the coefficient in a wavelet basis leads in this case to minimize exactly the same cost function as in the pixel case. Indeed

$$\sigma^2 \left(Wq - WI\right)^T \mathrm{I}_n \left(Wq - WI\right) = \sigma^2 (q-I)^T \mathrm{I}_n (q-I).$$

As a consequence, in §6.1 the difference between the pixel and wavelet operator is only due to the information selection through the thresholding operation.

Curvelet case: The curvelet transform by wrapping, defined in Candes et al. (2006), is a redundant (and therefore non invertible) transform. A pseudo–inverse of the curvelet transform by wrapping is the adjoint of the forward transform. It enables a perfect reconstruction of the original image given the full set of curvelet coefficients. The error covariance matrix in a curvelet space is nevertheless not invertible, and we do not try to construct it. As in the wavelet case, we rather work in a subspace defined through the thresholding $\mathcal{T}_I$ in (9). The covariance matrix is represented with a scalar:

$$R^{iid}_{\mathcal{T}_I(C),Scalar} = \hat{\sigma}^2 \mathrm{I}_k \tag{22}$$

where k stands for the number of coefficients kept by the thresolding operation.

Gradient case: The operator ∇ is not invertible. Neither is the error covariance matrix $C^{iid}_{\nabla} = \sigma^2 \nabla\nabla^T$. In section 6 we use two different approximations of this matrix:

- A scalar matrix corresponding to the diagonal of the true error covariance matrix:

$$R^{iid}_{\nabla,Scalar} = Diag(C^{iid}_{\nabla}) = \tilde{\sigma}^2 \mathrm{I}_{2n}. \tag{23}$$

In this study, as central differences are used to compute partial derivatives, the variance reads:

$$\tilde{\sigma}^2 = \frac{\sigma^2}{2}. \tag{24}$$

- A block diagonal matrix,

$$R^{iid}_{\nabla,Block} \approx C^{iid}_{\nabla} \tag{25}$$

built from C^{iid}_{∇} by making the hypothesis that there is no interaction between the error made on the x– and y– derivatives. The construction and inversion of this matrix is detailed in appendix 7.

Angular case: As the angular distance defined in §3.4 is not linear, it is difficult to build statistics for this case. Indeed, the formal computation of the mean and the covariances of the noise error are out of reach. Therefore, we consider that observations are only reliable when $\|\nabla I^o\| \gg \sigma$. Indeed in this case the noise does not have much impact on the gradient direction. Therefore we design the variance of both component of $\nabla I^o(i,j)/\|\nabla I^o(i,j)\|_\epsilon$ to be almost equal to σ^2 for $\|\nabla I^o\| \gg \sigma$, and very large for $\|\nabla I^o\| \ll \sigma$ (so that the observation is almost discarded). This writes as follows:

$$\overline{\sigma}^2_{i,j} = \frac{\sigma^2}{1-\exp\left(-\frac{||\nabla I^o||^2}{2\sigma^2}\right)}. \tag{26}$$

The covariance matrix used in twin experiments of section 6 is thus:

$$R^{iid}_{Ang,Diag} = \begin{pmatrix} R_x & 0 \\ 0 & R_y \end{pmatrix} \tag{27}$$

where we consider isotropic correlations through

$$R_x = R_y = \begin{pmatrix} \overline{\sigma}^2_{1,1} & 0 & \dots & \dots & 0 \\ 0 & \ddots & 0 & \cdots & \vdots \\ \vdots & \ddots & \overline{\sigma}^2_{i,j} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & \overline{\sigma}^2_{n,n} \end{pmatrix}. \tag{28}$$

5.3 Correlated gaussian white noise

Let us now assume that η is an additive Gaussian noise spatially correlated. The correlation is modeled by:

$$\eta = G \star \beta \tag{29}$$

where β is an independent and identically distributed gaussian white noise of variance σ, $\star$ is the convolution product symbol and G is a gaussian filter of size $(2n+1)(2n+1)$ such that

$$G(i,j) = \frac{1}{|G|} \exp\left(-\frac{i^2+j^2}{2\sigma_L^2}\right) \text{ for } i,j = -n,...,n. \tag{30}$$

where $|G| = \sum_{i,j} \exp\left(-\frac{i^2+j^2}{2\sigma_L^2}\right)$.

Pixel case: As we do not want to inverse a huge matrix C^{cor}_{Pix}, we only use information about variances. The noise is identically distributed so each pixel variance is the same. Therefore the approximation is scalar and reads:

$$R^{cor}_{Pix,Scalar} = Diag(C^{cor}_{Pix}) = \sigma^2 \mathrm{I}_n. \tag{31}$$

Wavelet case: One can build the error covariance matrix in wavelet space from the diagonal assumption in the pixel space given by equation (31). Doing this leads to define the approximation of the covariance matrix in any orthonormal wavelet subspace as:

$$\begin{aligned} R^{cor}_{W,Scalar} &= W Diag(C^{cor}_{pix}) W^T = \sigma^2 \mathrm{Id}_n \\ &\neq Diag(C^{cor}_W). \end{aligned} \tag{32}$$

The covariance matrix in a wavelet space C^{cor}_{Wav} can be theoretically constructed using equation (18) from C^{cor}_{Pix}. In Vannucci and Corradi (1999) the authors provide an algorithm based on the recursive discrete wavelet transform in one dimension. As we work with separable wavelets, it is easy to write a two–dimensional extension of this algorithm.

In this study we approximate the true error covariance matrix C^{cor}_W by its diagonal:

$$R^{cor}_{W,Diag} = \mathrm{Diag}(C^{cor}_W) = \mathrm{Diag}(W C^{cor}_{pix} W^T). \tag{33}$$

Such a process is applied with the Haar basis and the Daubechies basis with 8 vanishing moments. In the following, their covariances matrix are denoted as:

$$R^{cor}_{D8,Diag} \quad \text{for the Daubechies basis} \tag{34}$$

$$R^{cor}_{Haar,Diag} \quad \text{for the Haar basis} \tag{35}$$

Remark 4. When using the covariance matrix $R^{cor}_{Haar,Diag}$ or $R^{cor}_{D8,Diag}$ (defined in equation (33)-(35)) no thresh-

olding operation is applied on wavelet coefficients. As a consequence, the information considered is exactly the same in a pixel space or in those two spaces. The only difference resides in the error statistics specification.

One way to know what can be won (or lost) using such approximations is to compute their equivalence in the pixel space through:

$$\begin{aligned} R^{cor}_{Pix,Eq} &= W^{-1} R^{cor}_{W,Diag} W^{-T} \\ &= W^{-1} \mathrm{Diag}(W C^{cor}_{pix} W^T) W^{-T} \end{aligned} \tag{36}$$

Let us describe what kind of correlation in the pixel space are taken into account with the so–defined diagonal matrices in wavelet spaces. Figure 4 exhibits error covariances between a pixel and its neighbourhood for 9 pixels. On the top left panel, one can observe the true error covariances extracted from C^{cor}_{Pix}. The correlations are isotropic and the same for each pixel *far* from the boundary. On the top right, with a diagonal assumption in a pixel space, no spatial correlations are taken into account. On the bottom figures, error covariances in the pixel space induced by a diagonal assumption in Haar basis (left) and Daubechies basis (right) are displayed. Unlike the diagonal assumption in pixel space, those approximations take into account information on the spatial covariance. One can see that the correlation maps depends on the chosen wavelet space. We can notice that pixel variance induced by a diagonal assumption in Daubechies basis slightly differed from a pixel to another which is due to the fact that in a wavelet transform each pixel plays a specific role.

Gradient case: As the noise is identically distributed and the correlations are isotrope, the variance of each gradient component is the same. Taking into account exclusively information about variances leads to work with the scalar matrix:

$$R^{cor}_{\nabla,Scalar} = Diag(C^{cor}_{\nabla}) = \tilde{\sigma} \mathrm{I}_{2n}. \tag{37}$$

Note that, as central differences are used to compute partial derivatives, the variance reads:

$$\tilde{\sigma}^2 = \frac{\sigma^2 - p}{2} \tag{38}$$

where

$$p = \mathrm{cov}(I(i+1,j), I(i-1,j)) = \mathrm{cov}\left(I(i,j+1), I(i,j-1)\right). \tag{39}$$

Additionally a second test case is set up in order to measure the impact of a wrong matrix specification. This

second experiment involves the block matrix of equation (25) with a correlated noise. This matrix is not an approximation of the C_{∇}^{cor}:

$$R_{\nabla,Block}^{cor} = R_{\nabla,Block}^{iid} \neq Block(C_{\nabla}^{cor}). \tag{40}$$

Angular and curvelet case: As we do not know how to specify the effect of a correlated noise on those operators, we simply use the covariance matrix built for uncorrelated noise for those distances:

$$R_{\mathcal{T}_I(C)-Scalar}^{cor} = R_{\Omega_\tau^C-Scalar}^{iid} = \hat{\sigma}^2 \mathrm{I}_k,$$

$$R_{Ang,Diag}^{cor} = R_{Ang,Diag}^{iid}.$$

6 Twin experiments and robustness to noise

In this section the impact of data noise on the assimilation process is studied. In §6.1 we perform experiments with a gaussian space– and time– uncorrelated noise, then in §6.2 the impact of spatial correlation is discussed. For each kind of noise, different noise levels are tested. For each level, 10 noise scenarios are generated in order to perform 10 independent experiments.

In order to quantify the reliability of the information available in our image sequences, we refer to the Signal to Noise Ratio (SNR) which is computed as:

$$SNR = 10 \log_{10} \left(\frac{\sum_{\boldsymbol{x}\in\Omega} (I^*(\boldsymbol{x})^2)}{\sum_{\boldsymbol{x}\in\Omega} (I^*(\boldsymbol{x}) - I^o(\boldsymbol{x}))^2} \right) \tag{41}$$

which represents the logarithm of the ratio between Image Energy and Noise Energy. A large SNR ($\approx 30-40dB$) means that the image sequence has a good quality whereas a smaller (or negative) SNR means that the image sequence is strongly corrupted by noise. Figure 5 presents examples of images used in those twin experiments.

6.1 Independent Additive Gaussian white Noise

We first consider images corrupted by an additive Gaussian white noise uncorrelated in space and time. The noise is identically distributed with mean 0 and variance σ^2. We test the robustness of our operators to different noise scenarios, considering $\sigma \in [0; 0.1]$ (whereas $I^* \in [0.13; 0.86]$).

To do so, we quantify the improvement on the zonal velocity field u of the passive tracer field assimilation (which indirectly contains information on velocity). Table 2 presents the final values of the ratio $RMSE(u^a)/RMSE(u^b)$ w.r.t the SNR (see §4.1.3).

One can see that the *angular* operator is not well suited to deal with independent gaussian white noise. Indeed, even with images sequences of good quality, the error on the analyzed field remains important. This can be explained by the fact that the normalized gradient field holds less information than other distances. This is particularly true in flat areas where the information is not reliable, as taken into account by the covariance matrix modeling (26–28). On top of that, the angular covariance matrix modeling suffers from the fact that no correlations are considered. All those drawbacks could explain why the result obtained with this distance are worst than the results obtained with the other distances on the whole area (whereas they were the best with perfect data).

Curvelet ($\mathcal{T}_I(C) - Scalar$) and *wavelet* ($\mathcal{T}_I(W) - Scalar$) distances give estimations of the zonal velocity very close to the one induced by the pixel distance.

Results obtained with the block covariance matrix, $R^{iid}_{\nabla - Block}$ of equation (25), or the scalar matrix, $R^{iid}_{\nabla - Scalar}$ of equation (23) are very different. In this context, introducing some information on the correlation through the covariance matrix leads to a better analyzed field for all high noise levels.

To summarize, in case of additive uncorrelated gaussian white noise, all the observation operators are sensitive to the noise amplitude. In this particular case, we can not conclude that working in a space is better than working in another one. Indeed, results obtained in a *pixel*, a *wavelet*, a *curvelet* or even a *gradients* (if the covariance matrix is well specified) space are very close to each other.

6.2 Correlated Gaussian White Noise

Let us now consider an additive Gaussian noise spatially correlated as presented in section 5.3. In our study the isotropic correlation lenght of equation (30) is parameterized with $\sigma_L = 1.5$.

Table 3 presents the RMSE ratio of the zonal component of the velocity for different distance notions with respect to the SNR. Several important analyses can be drawn from these results:

- The best analyzed fields come from the distance in a wavelet basis combined with the true information on error variances ($D_8 - Diag$ or $Haar - Diag$). As all the coefficients are used for those distances, the information (including the noise) carried in a wavelet basis and the pixel basis is the same. Therefore, the difference between those distances and the $Pix - Scalar$ distance only lie in the way errors statistics are taken into account (see figure 4).
- One should notice that the $D_8 - Diag$ distance is more robust to noise amplitude than the $Haar - Diag$

distance. This comes from the fact that the correlation induced by a diagonal assumption in an Daubechies basis are more isotropic than the ones in Haar basis (see figure 4). Therfore it better fits the isotropy of the noise correlation. This illustrates that the choice of the wavelet basis should be related to the kind of observation noise in order to get a suitable approximation of the true covariance with a diagonal covariance matrix.

- Contrary to the uncorrelated case, $\nabla - Scalar$ distance outperforms $\nabla - Block$ distance. This difference is due to the use of a non valid approximation of the error covariance matrix in the second case. This experiment means that a good approximation in the uncorrelated case should not be applied without care in another case.
- As in the uncorrelated case $Pix - Scalar$ distance leads to results very close to $\mathcal{T}_I(C) - Scalar$ and $\mathcal{T}_I(D_8) - Scalar$ distances on the whole domain (table 3) and on a restricted domain (table 4). This is interesting as it demonstrates that the compression of the observations to a small set of coefficients (through the thresholding operation) conserves the pertinent information.
- $\nabla - Scalar$ distance outworks $Pix - Scalar$ distance. This result is due to the fact that gradients are less affected by a positively correlated noise (which can be seen as a local bias, which vanishes after differentiation) than pixels.
- We can also notice that the $D_8 - Diag$ distance using the worst images sequence (see figure 5 top right) gives better results than the other distances for all but one scenarios (even with the less corrupted image sequences presented in the bottom right of Figure 5). This illustrates that no matter the noise level, poor results are recovered if the error covariance matrix is not well defined. Such result demonstrates that the modeling of the error covariance matrix is a critical issue for assimilating images.

Table 4 presents the RMSE ratios computed for the sub-area presented in Figure 2. In this area, $Ang - Diag$ distance is competitive with other distances whereas it sytematically gives worst results on the whole domain. This behaviour is due to the fact that the *angular* observation operator is not a relevant tool in flat areas where the gradient direction is highly sensitive to noise, but gives pertinent information in the presence of image structures. As in the whole domain, $D_8 - Diag$ and $Haar - Diag$ distances present the best performance whatever the noise level.

Figure 6 presents the analysis error on v_0 for each observation operator considered. The results are presented for the worst noise level studied in a correlated case (see figure 5).

One should notice that for the *pixel* distance, the *curvelet* distance ($\mathcal{T}_I(C) - Scalar$), the *wavelet* distance

($\mathcal{T}_I(D8) - Scalar$), and the *gradients* distance ($\nabla - Block$) the error presents the same structure. The only difference lies in the amplitude of each structure. This phenomenon is due to the fact that the information and the errors statistics used for these distances are nearly the same.

Althougth the level of error in the analyzed state obtained with the *angular* distance is nearly the same as the previous operators (see table 3 for the u–component of the velocity), the analysis error distribution differs a lot from the previous ones. This is due to the fact that in this space another kind of information is considered.

Compared to the previous distance the analysis obtained with the *wavelet* distances associated to a diagonal covariance matrix or the *gradients* distance associated to a scalar matrix are less corrupted by noise. This is particularly true outside the vortex area where the information on error correlation incorporated in a wavelet space play an important role. In this region, as the true tracer concentration is nearly a constant, the image does not bring any piece of information on the underlying dynamic to constrain the data assimilation system. As a consequence, in this area, the correlation coefficients control the way the analysis can differ from the noisy observation field.

These experiments have highlighted that, for spatially correlated observation errors, the $D_8 - Diag$ distance performs better than others. Indeed the diagonal matrix associated to this distance implicitly takes into account, for this kind of noise, an appropriate approximation of the noise spatial correlations, without requiring computational expensive matrix inversions.

Among the scalar covariance matrices (mostly used in operational DA systems) the $\nabla - Scalar$ distance presents the best analyzed field. This comes from the fact that gradients coefficient are less affected by a noise involving positive correlation (which can be seen as a local bias) than the coefficients in other spaces.

7 Conclusion

This paper first presents two original observation operators and associated norms dedicated to image–type observation assimilation. Both are based on the gradient of the image, in norm and orientation respectively. They showed competitive performances on an academic test case compared to previously introduced metrics and therefore are good candidates for more demanding applications. In the future, it could be interesting to study combinations of such metrics in order to benefit from their respective merits.

The second part of this paper highlights the importance of a good specification of the observation error covariance

matrix in data assimilation. Indeed this is a crucial point for image–type data, but is likely to be of significant importance for other kind of data. This was illustrated thanks to the study of two kind of observation noise. First an uncorrelated additive Gaussian noise, all operators (classical and new) show similar robustness (except for the *angular* operator which suffers from the variance specification) to the noise level, and none of them seems any worse or better. This is a favourable case that is seldom encountered in practice. More realistically, a second set of experiment is performed applying a correlated additive Gaussian noise to the data. In that case, it appears clearly that it is important to improve the observation error covariance matrix specification and avoid the use of scalar ones (*i.e.* $R = \sigma I$), as it is commonly done in practice. Interestingly, such a process can be done at a cheap cost by considering diagonal covariance matrix in other space (for example a wavelet space). It is also pointed out that the specification of the covariance matrices should obviously be related to the actual observation noise and that it is not just a matter of using a non–diagonal R.

This last remark means that a careful study of the observation error has to be done prior to assimilation. In this paper we assumed here that R is known in the pixels space, but in practice, how could we adapt this work to cases where R should be modelled from scratch? On that point, a wide literature exist in different field such as images and statistics which aims to estimate and/or model the variance terms in a different space, we can for example mention the interesting work of Nguyen van Yen et al. (2012) which iteratively estimate the wavelet variance scale by scale. In variational data assimilation, the balance between the background and observation error matrices B and R is a subtle alchemy, improving the R description should allow for a more consistent B specification.

Another crucial point to sort out prior to operational use, specific to image–type observation, is to address the problem of occultation (a cloud passing by over ocean surface data, for instance) and its impact on error statistics. Again, wavelet description, or combination of metrics presented in this paper, may offer a suitable framework for dealing with such problem.

Acknowledgements

This work got support from the LEFE program of INSU (CNRS) and from the french national research agency project Geo–FLUIDS n°ANR–09–COSI–005.

APPENDIX: Covariance matrix for gradient additive noise

1 Block approximation

Recalling that the observation operators reads: $\nabla I = \nabla q + \nabla \eta$. A possible discrete implementation of the gradient $\nabla \eta = [\eta_x, \eta_y]^T$ at pixel (i,j) with $i = 1 \cdots L$ and $j = 1 \cdots H$, is defined as:

$$\eta_x(i,j) = \frac{\eta(i,j+1) - \eta(i,j-1)}{2} \text{ if } 1 < j < H \tag{1}$$

and

$$\eta_y(i,j) = \frac{\eta(i+1,j) - \eta(i-1,j)}{2} \text{ if } 1 < i < L \tag{2}$$

Even if the noise is a white Gaussian additive noise, there are some correlation between the gradient components. Indeed, when $\eta_x(i,j)$ is inside the domain boundary, it is correlated with $\eta_x(i+2,j)$, $\eta_x(i-2,j)$, $\eta_y(i-1,j-1)$, $\eta_y(i+1,j-1)$, $\eta_y(i-1,j+1)$, $\eta_y(i,j-1)$.

As illustrated in figure 7 the covariance coefficients are:

$$\begin{aligned} cov\left(\eta_x(i,j), \eta_y(i-1,j+1)\right) &= cov\left(\eta_x(i,j), \eta_y(i+1,j-1)\right) \\ &= \frac{\sigma^2}{4} \end{aligned}$$

$$\begin{aligned} cov\left(\eta_x(i,j), \eta_x(i,j+2)\right) &= cov\left(\eta_x(i,j), \eta_x(i,j-2)\right) \\ &= cov\left(\eta_x(i,j), \eta_y(i-1,j-1)\right) \\ &= cov\left(\eta_x(i,j), \eta_y(i+1,j+1)\right) \\ &= -\frac{\sigma^2}{4} \end{aligned}$$

which leads to a non invertible R matrix written as:

$$R = \begin{pmatrix} C_{\eta_x\eta_x} & C_{\eta_x\eta_y} \\ C_{\eta_y\eta_x} & C_{\eta_y\eta_y}. \end{pmatrix}$$

where $C_{\eta_x\eta_x}$ is the error covariance matrix between the x– derivative elements. Simplifying this covariance matrix by neglecting the correlation between x– and y– derivative elements leads to consider that $C_{\eta_x,\eta_y} = 0$. This simplification is illustrated on figure 8. One consequence is that one element η_x (resp. η_y) is only correlated to elements of the same row (resp. column).

This approximation leads to consider the following block matrix $\tilde{R}$:

$$\tilde{R} = \begin{pmatrix} C_{\eta_x^{i=1}\eta_x^{i=1}} & 0 & \cdots & 0 \\ 0 & C_{\eta_x^{i=2}\eta_x^{i=2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & C_{\eta_y^{j=N}\eta_y^{j=N}} \end{pmatrix}$$

where $C_{\eta_x^{i=1}\eta_x^{i=1}}$ is the block corresponding to the covariance between $x-$ derivative elements of the first row.

2 Inversion of $\tilde{R}$

Albeit the boundary elements, a block $C_{\eta_x^{i=k}\eta_x^{i=k}}$, of the matrix $\tilde{R}$ can be separated in two blocks depending on j parity:

$$C_{\eta_x^{i=k}\eta_x^{i=k}} = \begin{pmatrix} C_{\eta_x^{j=odd}\eta_x^{j=odd}} & 0 \\ 0 & C_{\eta_x^{j=even}\eta_x^{j=even}} \end{pmatrix}$$

Taking into account that all blocks are identical we only need to invert four tridiagonal matrix $C_{\eta_x^{odd}\eta_x^{odd}}$, $C_{\eta_x^{even}\eta_x^{even}}$, $C_{\eta_y^{odd}\eta_y^{odd}}$, $C_{\eta_y^{even}\eta_y^{even}}$.

If $L = H = 2N$, only one block matrix A (which size is $(N-1)\times(N-1)$) needs to be inverted to fill–in the full covariance matrix $\tilde{R}$. This matrix reads:

$$A = \sigma^2 \begin{pmatrix} 0.5 & -0.25 & 0 & \cdots & 0 \\ -0.25 & 0.5 & -0.25 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & 0 & -0.25 & 0.5 & -0.25 \\ \vdots & \vdots & 0 & -0.25 & 0.5 \end{pmatrix} \tag{3}$$

Its inverse exists and reads as:

$$A^{-1} = \frac{4}{N\sigma^2} D$$

with

$$D = \begin{pmatrix} N-1 & N-2 & N-3 & \cdots \\ N-2 & 2*(N-2) & 2*(N-3) & \cdots \\ N-3 & 2*(N-3) & 3*(N-3) & \cdots \\ \vdots & \vdots & \vdots & \ddots \\ 1 & 2 & \cdots & N-1 \end{pmatrix}$$

REFERENCES

L. Aouf, J.-M. Lefèvre, and D. Hauser. Assimilation of directional wave spectra in the wave model wam: An impact study from synthetic observations in preparation for the swimsat satellite missionl. *Journal of Atmospheric and Oceanic Technology*, 23: 448–463, 2006.

S. Ba and R. Fablet. Variational fronts tracking in sea surface temperature images. In *IEEE International Conference on Image Processing (ICIP'10)*, pages 4117 –4120, 2010.

J. L. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of optical flow techniques. *International journal of computer vision*, 12(1):43–77, 1994.

N. Bormann, S. Saarinen, and G. Kelly. The spatial structure of observation errors in atmospheric motion vectors from geostationary satellite data. *Mon. Wea. Rev.*, 131(4):706–718, 2003.

E. Candes, L. Demanet, D. Donoho, and L. Ying. Fast discrete curvelet transforms. *Multiscale Modeling & Simulation*, 5(3):861–899, 2006.

C. Cardinali. Monitoring the observation impact on the short-range forecast. *Quarterly Journal of the Royal Meteorological Society*, 135(638):239–250, 2009. ISSN 1477-870X. 10.1002/qj.366. URL http://dx.doi.org/10.1002/qj.366.

T. Corpetti, P. Héas, E. Mémin, and N. Papadakis. Pressure image assimilation for atmospheric motion estimation. *Tellus Series A: Dynamic Meteorology and Oceanography*, 61(1):160–178, 2009.

G. Desroziers, L. Berre, B. Chapnik, and P. Poli. Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society*, 131(613):3385–3396, 2005.

G. Desroziers, L. Berre, V. Chabot, and B. Chapnik. A posteriori diagnostics in an ensemble of perturbed analyses. *Monthly Weather Review*, 137:3420–3436, 2009.

F. D'Ovidio, V. Taillandier, L. Mortier, and I. Taupier-Letage. Lagrangian validation of the mediterranean mean dynamic topography by extraction of tracer frontal structures. *Mercator Ocean Quarterly Newsletter*, 32:24–32, 2009.

J. Fadili, J.-L. Starck, M. Elad, D. Donoho, et al. Mcalab: Reproducible research in signal and image decomposition and inpainting. *IEEE Computing in Science and Engineering*, 12(1):44–63, 2010.

J.-B. Flór and I. Eames. Dynamics of monopolar vortices on a topographic beta-plane. *Journal of Fluid Mechanics*, 456(1):353–376, 2002.

L. Gaultier, J. Verron, J.-M. Brankart, O. Titaud, and P. Brasseur. On the inversion of submesoscale tracer fields to estimate the surface ocean circulation. *Journal of Marine Systems*, 2012.

J. C. Gilbert and C. Lemaréchal. Some numerical experiments with variable-storage quasi-newton algorithms. *Mathematical programming*, 45(1-3):407–435, 1989.

B. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185203, 1981.

G. Korotaev, E. Huot, F.-X. Le Dimet, I. Herlin, S. Stanichny, D. Solovyev, and L. Wu. Retrieving ocean surface current by 4-d variational assimilation of sea surface temperature images. *Remote Sensing of Environment*, 112(4):1464 – 1475, 2008.

F.-X. Le Dimet and O. Talagrand. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus*, pages 97 – 110, 1986.

B. Lucas and T. Kanade. An iterative image registration technique with an application to stereovision. In *Int. Joint Conf. on Artificial Intel. (IJCAI'81)*, pages 674–679, 1981.

Y. Michel. Displacing potential vorticity structures by the assimilation of pseudo-observations. *Monthly Weather Review*, 0(0):null, 2011.

R. Nguyen van Yen, M. Farge, and K. Schneider. Scale-wise coherent vorticity extraction for conditional statistical modeling of homogeneous isotropic two-dimensional turbulence. *Physica D Nonlinear Phenomena*, 241:186–201, 2012.

S. Nieman, P. Menzel, C. Hayden, D. Gray, S. Wanzong, C. Velden, and J. Daniels. Fully automated cloud-drift winds in NESDIS operations. *Bull. Amer. Meteor. Soc*, 78:1121–1133, 1997.

O. Pannekoucke. Heterogeneous correlation modeling based on the wavelet diagonal assumption and on the diffusion operator. *Monthly Weather Review*, 137(9):2995–3012, 2009.

N. Papadakis and E. Mémin. A variational framework for spatio-temporal smoothing of fluid motions. In *Scale Space and Variational Methods, SSVM'07*, pages 603–615, May 2007.

N. Papadakis and E. Mémin. Variational assimilation of fluid motion from image sequence. *SIAM Journal on Imaging Sciences*, 1: 343–363, 2008.

J. Schmetz, K. Holmlund, J. Hoffman, B. Strauss, B. Mason, V. Gaertner, A. Koch, and B. van De. Operational cloud-motion winds from Meteosat infrared images. *Journal of applied meteorology*, 32(7):1206–1225, 1993.

I. Souopgui. *Assimilation d'images pour les fluides géophysiques*. PhD thesis, Université de Grenoble, 2010.

L. M. Stewart, S. Dance, and N. K. Nichols. Data assimilation with correlated observation errors: experiments with a 1-d shallow water model. *Tellus A*, 65, 2013.

O. Titaud, A. Vidard, I. Souopgui, and F.-X. Le Dimet. Assimilation of image sequences in numerical models. *Tellus Series A: Dynamic Meteorology and Oceanography*, 62(1):30–47, 2010.

O. Titaud, J.-M. Brankart, and J. Verron. On the use of finite-time lyapunov exponents and vectors for direct assimilation of tracer images into ocean models. *Tellus Series A: Dynamic Meteorology and Oceanography*, 63(5):1038–1051, 2011.

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.

A. Weaver and P. Courtier. Correlation modelling on the sphere using a generalized diffusion equation. *Quarterly Journal of the Royal Meteorological Society*, 127:1815–1846, July 2001.

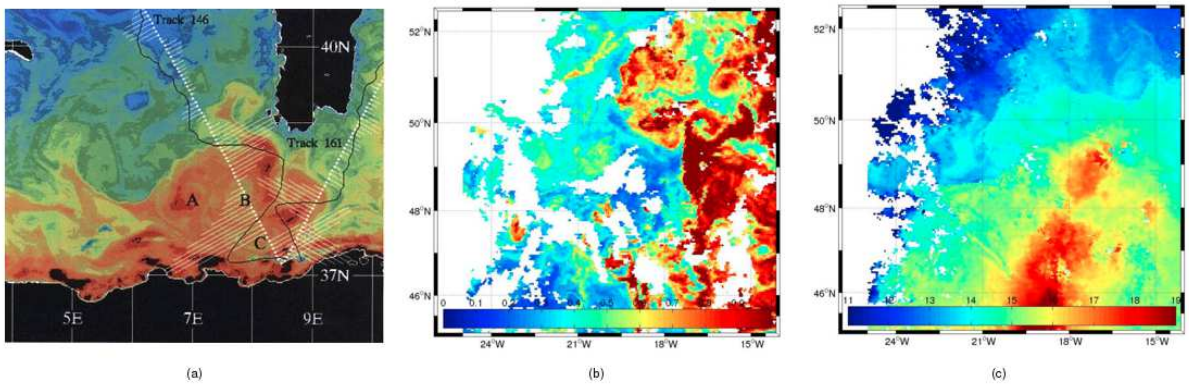

Figure 1. Different examples of satellite images. (a) Altimetric reconstruction from JASON satellite data. (b) Ocean color/ Chlorophyllis from the MODIS captor of ENVISAT satellite. (c) Sea surface Temperature from the MODIS captor of ENVISAT satellite.

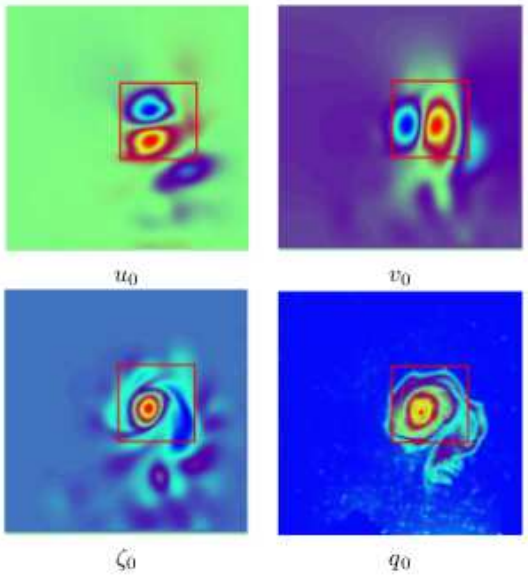

Figure 2. Initial values of the scenario taken for the twin experiments. The zonal and meridional velocities u_0^* and v_0^* are characterized by a strong vortex, illustrated by the initial vorticity ζ_0. The height h_0^* is assumed almost flat. The velocities take their values from $-0.03ms^{-1}$ (blue) to $0.04ms^{-1}$ (red) and the 2D vorticity from $-0.2s^{-1}$ (blue) to $0.64s^{-1}$ (red). This synthetic initialization has been created in order to match correctly with the initialization of the passive tracer, q_0^*. The red box defines the subdomain we will use to compute the RMS, when we want to focus the method assessment on the vortex area.

Table 1. Ratio between analyzed state RMS and background state RMS for perfect data. This ratio is computed over the whole domain starting from a background at rest.

	$\frac{RMSE(u^a)}{RMSE(u^b)}$	$\frac{RMSE(v^a)}{RMSE(v^b)}$	$\frac{RMSE(\zeta^a)}{RMSE(\zeta^b)}$	$\frac{RMSE(\alpha^a)}{RMSE(\alpha^b)}$
Curvelet	4.8%	3.6%	11.2%	27.9%
Wavelet	4.3%	3.3%	10.2%	23.4%
Pixels	4.7%	3.5%	11.4%	24.5%
Angular	3.6%	2.5%	9.1%	21.5%
Gradient	4.8%	3.5%	11.0%	23.3%

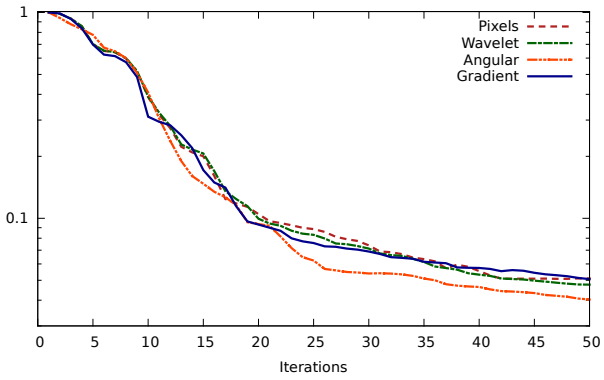

Figure 3. Evolution of the ratio of RMS errors of the velocity u with respect to 4D–Var iterations. Perfect data are observed and the four observators are compared.

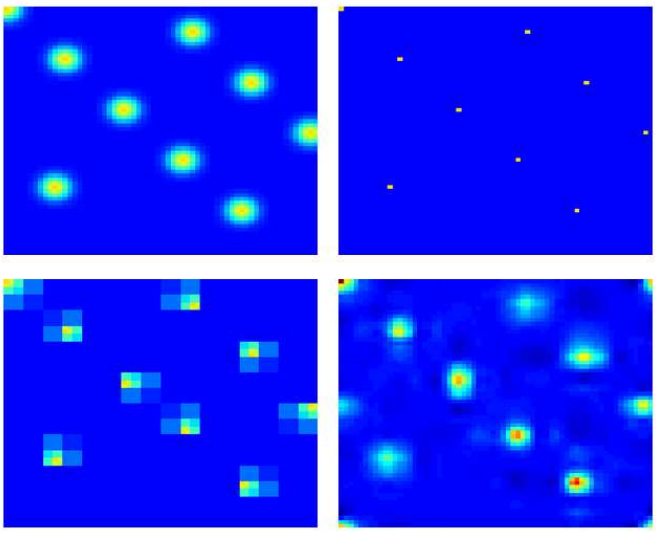

Figure 4. Top left : True errors covariances with their neighbourhood for 9 pixels. Top right : Diagonal approximation in a Pixel space. Bottom : Errors covariances in the pixel space equivalent to a diagonal assumption for Haar wavelets (left) and for Daubechies with 8 vanishing moments wavelets (right) for the same pixels.

Table 2. Ratio between analysed state RMSE and background state RMSE for data with additive iid noise of decreasing standard deviation. The variable of interest is u. The RMSE are computed on the whole domain.

SNR	6.8	12.3	26.2	46.3
$\mathcal{T}_I(C) - Scalar$	76.4%	37.1%	13.0%	6.1%
$\mathcal{T}_I(W) - Scalar$	78.5%	35.9%	12.3%	5.5%
$Pix - Scalar$	61.7%	32.0%	11.1%	5.3%
$\nabla - Block$	62.7%	28.7%	10.8%	5.1%
$\nabla - Scalar$	Divergence	48.8%	12.3%	4.7%
$Ang - Diag$	80.3%	62.2%	31.1%	10.9%

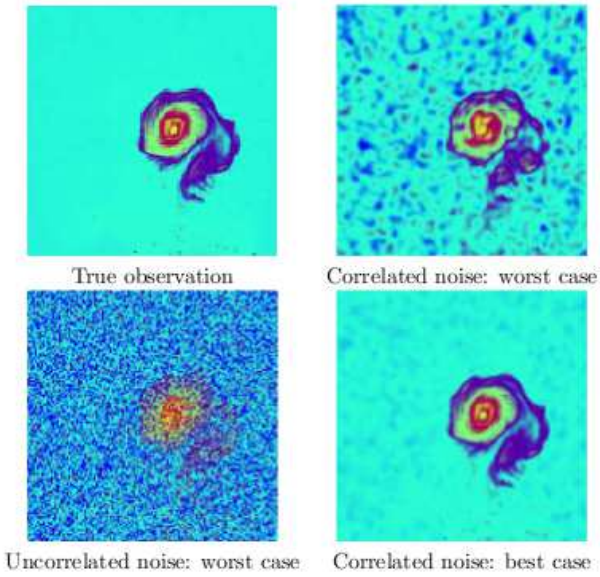

Figure 5. First image of the sequence used for the assimilation. The tracer concentration are plot from 0 (blue) to 1 (red). Top left: Perfect data scenario. Top right: worst scenario tested for additive correlated gaussian white noise (SNR = 14.8 dB). Bottom left: worst scenario for gaussian white noise (SNR = 6.7 dB). Bottom right: best scenario for an additive correlated white noise (SNR = 26.8 dB).

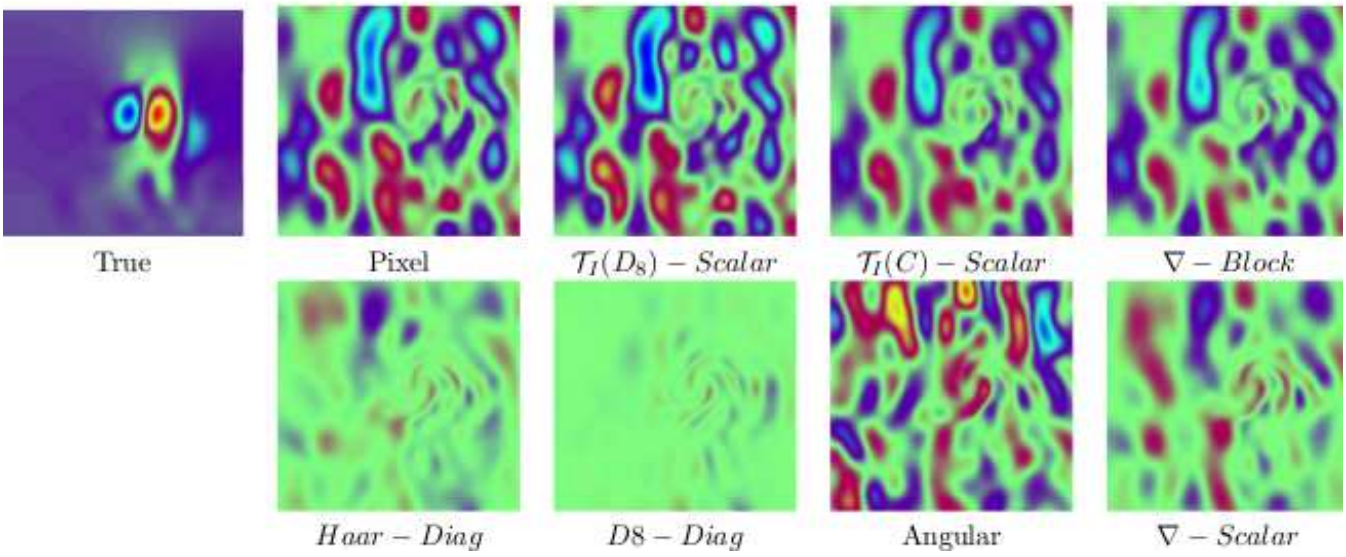

Figure 6. Error analysis on the v–component of the velocity using the worst correlated image sequence studied. The velocity errors are plot from $-0.010ms^{-1}$ (blue) to $0.010ms^{-1}$ whereas the true field is plot for $-0.025ms^{-1}$ (blue) to $0.040ms^{-1}$.

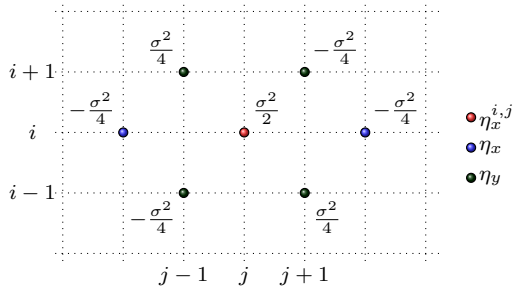

Figure 7. Covariance between $\eta_x(i,j)$ (in red) and other elements (in green and blue) in a gradient space for an independent and identicaly distributed gaussian white noise in the pixel space. The blue (resp. the green) points represents x– (resp. y–) derivative elements correlated to $\eta_x(i,j)$. All the covariances between $\eta_x(i,j)$ and other points are represented.

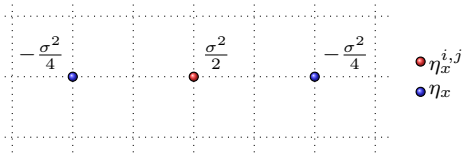

Figure 8. Illustration of the simplification of the error covariance matrix. Only correlation between derivative in the same direction are considered.

Table 3. Ratio between analyzed state RMSE and background state RMSE for data with spatially correlated additive noise of decreasing standard deviation. The variable of interest is u. The RMSE are computed on the whole domain.

SNR	14.8	20.8	26.8
$\mathcal{T}_I(C) - Scalar$	66.8%	29.4%	18.6%
$\mathcal{T}_I(D_8) - Scalar$	64.3%	26.0%	15.2%
$Haar - Diag$	22.5%	12.3%	8.2%
$D_8 - Diag$	9.1%	7.5%	7.0%
$Pix - Scalar$	59.8%	25.6%	15.1%
$\nabla - Block$	62.7%	28.7%	15.8%
$\nabla - Scalar$	33.3%	17.3%	11.8%
$Ang - Diag$	59.8%	42.6%	31.5%

Table 4. Ratio between analysed state RMSE and background state RMSE for data with spatially correlated additive noise of decreasing standard deviation. The variable of interest is u. The RMSE are only computed on a domain centered around the vortex (see the red box on figure 2) which contain most of image structure information.

SNR	14.8	20.8	26.8
$\mathcal{T}_I(C) - Scalar$	20.1%	9.9%	8.0%
$\mathcal{T}_I(D_8) - Scalar$	19.7%	8.7%	6.5%
$Haar - Diag$	7.0%	5.3%	4.6%
$D_8 - Diag$	5.0%	5.0%	4.8%
$Pix - Scalar$	19.7%	9.3%	7.5%
$\nabla - Block$	18.5%	9.4%	7.2%
$\nabla - Scalar$	10.7%	6.8%	5.6%
$Ang - Diag$	13.8%	9.8%	6.8%