Ensemble Kalman Filter based on the image structures

Dominique Béréziat, Isabelle Herlin, Yann Lepoittevin

Ensemble Kalman Filter based on the image structures

Dominique Béréziat[1], Isabelle Herlin[2] and Yann Lepoittevin[2]
[1]*Sorbonne Universités, UPMC Univ Paris 06, CNRS, LIP6 UMR 7606*
[2]*Inria, Paris*
Dominique.Bereziat@lip6.fr,lepoittevin.yann@gmail.com,Isabelle.Herlin@inria.fr

Keywords:
Domain Decomposition, Ensemble Kalman Filter, Fluid Flows, Localization, Motion Estimation.

Abstract:
One major limitation of the motion estimation methods that are available in the literature concerns the availability of the uncertainty on the result. This is however assessed by a number of filtering methods, such as the ensemble Kalman filter (EnKF). The paper consequently discusses the use of a description of the displayed structures in an ensemble Kalman filter, which is applied for estimating motion on image acquisitions. An example of such structure is a cloud on meteorological satellite acquisitions. Compared to the Kalman filter, EnKF does not require propagating in time the error covariance matrix associated to the estimation, resulting in reduced computational requirements. However, EnKF is also known for exhibiting a shrinking effect when taking into account the observations on the studied system at the analysis step. Methods are available in the literature for correcting this shrinking effect, but they do not involve the spatial content of images and more specifically the structures that are displayed on the images. Two solutions are described and compared in the paper, which are first a dedicated localization function and second an adaptive domain decomposition. Both methods proved being well suited for fluid flows images, but only the domain decomposition is suitable for an operational setting. In the paper, the two methods are applied on synthetic data and on satellite images of the atmosphere, and the results are displayed and evaluated.

1 Introduction

The paper describes the use of filtering methods for an online estimation of motion, associated with its uncertainty measure, on image sequences.

Dense motion estimation approaches originate from the classic optical flow algorithm (Horn and Schunk, 1981): the velocity field is computed as the minimum of an energy function, whose main assumption is the conservation of image brightness on each point trajectory. This method has been intensively studied and improved over time and gave birth to a huge number of papers in the literature. It still faces the issue of the uncertainty on the result.

Approaches such as the Kalman filter (Kalman, 1960) were introduced in this context for estimating the uncertainty on the motion result. The Kalman filter uses an analytical description of the Probability Density Function associated to the error. It improves the estimation at each measurement (image acquisition) as it updates simultaneously the state vector and its associated covariance matrix. This filter is directly used in (Elad and Feuer, 1998) on the image brightness values, but with images on size 50×50. The Kalman filter is also applied by (Franke and Rabe, 2005) and (Franke et al., 2005) on the KLT-tracker features described in (Tomasi and Kanade, 1991). As the result is a sparse motion field, an extension is defined in (Rabe et al., 2010), where a Kalman filter is defined for each pixel of the image domain, resulting in a dense motion estimation. However, as each filter is defined on a single point, and not on the whole domain, it does not include the correlations between two different pixels of the domain. Defining a global filter, whose state vector includes a number of variables for each pixel, is not directly usable for large size images, due to the huge dimension of the corresponding covariance matrix (squared number of pixels of

the image) that must be propagated in time.

For operational use, the solution for estimating the uncertainty on large size systems comes from the ensemble Kalman filter (EnKF) described in (Evensen, 2003). The approach relies on a sample of the Probability Density Function associated to the error from an ensemble of state vectors describing the system.

In case of motion estimation from an image sequence, an ensemble of motion and image fields is designed at initial time, integrated in time and sequentially improved from the observational image acquisitions. We proposed and described in (Lepoittevin et al., 2015) the construction of the initial ensemble. A set of state-of-the-art optical flow algorithms (around 50 methods, described in the surveys (Sun et al., 2010) and (Baker et al., 2011)) are used to design the initial ensemble of motion fields. The HS-brightness method (Sun et al., 2010) is one of these algorithms and it provides the best performance, when used alone, on our synthetic data. HS-brightness is a specific implementation of the Horn-Schunk algorithm (Horn and Schunk, 1981), which includes a multi-resolution scheme and a median filtering (Sun et al., 2010). Each member of the ensemble is then integrated in time with the model and an estimate of motion is computed, at each time, as the average of the motion members. The uncertainty is described by the spread of that ensemble. As in (Lepoittevin et al., 2015) the paper concerns the application of EnKF for estimating a dense motion field, based on the structures displayed by images. As the application domain concern fluid flows images, we do not rely on object characteristics, such as the SIFT features (Lowe, 1999) and variants, since they are not significant on these data. The innovation of the paper, compared to (Lepoittevin et al., 2015), concerns the use of a segmentation of images by EnKF. Moreover, we will stress, from the experiments, that merging a number of optical flow algorithms gives better results than the best one alone (the HS-brightness method): all methods take part to the result of EnKF.

Two different alternatives for characterizing the structures, which are displayed on fluid flows images, such as fronts and vortices on ocean satellite data, are compared in the context of EnKF. Comparison concerns two criteria: quality of results and computational performances.

The first approach, introduced in (Lepoittevin et al., 2015), concerns the design of a localization function that includes information on the displayed structures, as it has been discussed for instance in (Anderson, J. L., 2001), (Houtekamer and Mitchell, 1998), (Hamill et al., 2001), (Oke et al., 2007) and (Anderson, 2007).

The second approach applies a domain decomposition technique, as in (Nerger et al., 2006) and (Hunt et al., 2007), which depends on the image brightness values. This domain decomposition is equivalent to a segmentation of the image acquisitions.

Both techniques make the estimation depending on the structures and on their evolution in time.

As the Kalman filter itself is well-known in the image processing community, Section 2 will shortly summarize that point and only discuss the mathematical equations of the ensemble Kalman filter, as initially given in (Evensen, 2003). Section 3 reminds about the design of the localization function from image brightness values and its use in the EnKF formalism. The resulting method is named Explicit Structures Localization in the remaining of the paper. Section 4 explains the use of domain decomposition in the context of EnKF and describes the domain decomposition associated to a segmentation process. Results are given in Section 5, which compares the two approaches. The paper ends with some conclusions and hints on future research work.

2 EnKF and explicit localization

Let us first provide the notations that are further required in the paper.

Images are acquired on the spatial domain Ω. A pixel is denoted $\mathbf{p}$.

A sequence of $(\mathrm{N}_O + 1)$ acquisitions A_t, $\{t \in [0, \mathrm{N}_O]\}$ is processed. An observation vector $\mathbf{Y}_t$ is computed on each acquisition A_t. This vector is either a representation, line by line, of the image or a description of quantities that are computing on this image.

A state vector $\mathbf{X}$ of size $\mathrm{N}_{\mathbf{X}}$ is defined, whose value at time t is $\mathbf{X}_t = (\mathbf{w}_t, I_t)^T$. $\mathbf{w}_t$ is the vector describing the motion field and I_t is the vector associated with a synthetic image. The assumption is that the synthetic image sequence corresponding to I_t satisfies the optical flow equation: the image brightness is advected by the motion field.

The objective is to get an estimate $\mathbf{X}_t^{(a)}$ of the true state vector (and consequently of the true motion field), from its background value $\mathbf{X}_t^{(b)}$ and from the observation vector $\mathbf{Y}_t$, so that I_t is as

close as possible to $\mathbf{Y}_t$. The background value $\mathbf{X}_t^{(b)}$ is obtained from time integration of the estimation $\mathbf{X}_{t-1}^{(a)}$ at previous time.

Ensemble methods rely on a number N_m of members. $\mathbf{X}_t^i$ denotes the state vector at time t of the i^{th} member of the ensemble. The average $\mathbf{X}_t$ of the state vectors $\mathbf{X}_t^i$ is defined by:

$$\mathbf{X}_t = \frac{1}{\mathrm{N}_m} \sum_{i=1}^{\mathrm{N}_m} \mathbf{X}_t^i = \overline{\mathbf{X}_t^i} \tag{1}$$

Error terms are discussed in the paper according to the following notation: $\mathcal{E}_R$ is a centered Gaussian noise associated to the covariance matrix R and denoted:

$$\mathcal{E}_R \sim \mathcal{N}(0, R) \tag{2}$$

The equations of the Kalman filter (Kalman, 1960) are the following:

1. Let $\mathbb{M}$ be the numerical model describing the state vector dynamics. Readers should remind that the studied dynamics is non linear when studying fluid flows, as advection and convection processes are involved in the evolution. The background value $\mathbf{X}_t^{(b)}$ is obtained from the integration with $\mathbb{M}$ of the estimation $\mathbf{X}_{t-1}^{(a)}$, computed at $(t-1)$:

$$\mathbf{X}_t^{(b)} = \mathbb{M}\mathbf{X}_{t-1}^{(a)} \tag{3}$$

   The propagation in time of the uncertainty covariance matrix satisfies:

$$B_t^{(b)} = \mathbb{M}B_{t-1}^{(a)}\mathbb{M}^T \tag{4}$$

2. If no observation is available at t, the estimation $\mathbf{X}_t^{(a)}$ and the matrix $B_t^{(a)}$ are taken equal to those of the background.

3. If an observation vector $\mathbf{Y}_t$ is available at t, the estimation is computed by:

$$\mathbf{X}_t^{(a)} = \mathbf{X}_t^{(b)} + K\left(\mathbf{Y}_t - \mathbb{H}\mathbf{X}_t^{(b)}\right) \tag{5}$$

   $\mathbb{H}$ is the operator that projects the state vector $\mathbf{X}_t$ in the vector space of $\mathbf{Y}_t$. K is the Kalman gain, defined by:

$$K = B_t^{(b)}\mathbb{H}^T\left(\mathbb{H}B_t^{(b)}\mathbb{H}^T + R_t\right)^{-1} \tag{6}$$

   where R_t is the covariance matrix associated to the observation vector. The covariance matrix of the estimation error is equal to:

$$B_t^{(a)} = B_t^{(b)} - K\mathbb{H}B_t^{(b)} \tag{7}$$

The use of EnKF to replace the Kalman filter comes from an analysis of the computational requirements. The Kalman filter requires to store the whole analytical matrix $B_t^{(b)}$, whose size $\mathrm{N}_{\mathbf{X}}^2$ is usually huge when processing images: almost 10^7 for a 512×512 pixels image with 3 values at each pixel (2 for motion and 1 for brightness). Moreover, the propagation in time of this matrix relies on Equation (4), with a cost $\mathrm{N}_{\mathbf{X}}^3$. When applying EnKF, an ensemble of state vectors $\mathbf{X}_t^{(b),i}$ is defined at each time t and an approximation of the covariance matrix is obtained as:

$$B_t^{(b)} \approx \overline{(\mathbf{X}_t^{(b),i} - \mathbf{X}_t^{(b)})(\mathbf{X}_t^{(b),i} - \mathbf{X}_t^{(b)})^T} \tag{8}$$

with $\mathbf{X}_t^{(b)} = \overline{\mathbf{X}_t^{(b),i}}$ the average of the ensemble members. The whole knowledge available on the system is then included in the $\mathrm{N}_m\mathrm{N}_{\mathbf{X}}$ matrix containing all members. The number of members, N_m, being usually less than one hundred, the storage cost is drastically reduced.

Initialized at time 0, the ensemble is integrated in time by the model $\mathbb{M}$ in Equation (3). If no observation is available at time t, the estimation $\mathbf{X}_t^{(a),i}$ is equal to $\mathbf{X}_t^{(b),i}$ and the uncertainty is approximated with Equation (8). If an observation $\mathbf{Y}_t$ is available, $\mathbf{X}_t^{(a),i}$ is computed according to Equations (5) and (6). The estimation provided by the ensemble at time t is calculated as the average of members:

$$\mathbf{X}_t^{(a)} = \overline{\mathbf{X}_t^{(a),i}} = \frac{1}{\mathrm{N}_m} \sum_{i=1}^{\mathrm{N}_m} \mathbf{X}_t^{(a),i} \tag{9}$$

Its uncertainty is approximated by replacing $^{(b)}$ by $^{(a)}$ in Equation (8).

One important remark on the implementation of EnKF is that all members are involved in the computation of $B_t^{(b)}$, in Equation (8). Therefore, all members are included in the computation of the Kalman gain in Equation (6). Consequently, all members impact the estimation of member i, which is $\mathbf{X}_t^{(a),i}$ in Equation (5). The members are consequently depending one from the others, when evolving in time. We will come back to that issue when analyzing results.

EnKF is attractive for both its reduced computational requirements and its ability to describe the uncertainty of the result by the members. However, it also suffers from limitations. One major weakness of this filter is the approximate knowledge of the background covariance matrix $B_t^{(b)}$ obtained with Equation (8). This approximation leads to the appearance of spurious covariance values in the matrix, which impact the

quality of the estimation. The next two sections describe solutions to this limitation, which both rely on the structures displayed on the fluid flows images.

3 Structured Explicit Localization

In this section, we remind the method introduced in (Lepoittevin et al., 2015). This method is an explicit localization method. It corrects the background error covariance matrix $B_t^{(b)}$, in order to recover most of its analytical properties. For that purpose, the matrix $B_t^{(b)}$ is multiplied by a localization matrix ρ before being used in Equation (6):

$$L_t^{(b)} = \rho \circ B_t^{(b)} \tag{10}$$

with $\circ$ the point-wise matrix product.

The matrix $L_t^{(b)}$ is then included in place of $B_t^{(b)}$ in the equations that compute the analysis $\mathbf{X}_t^{(a),i}$ for the member i at time t:

$$\mathbf{X}_t^{(a),i} = \mathbf{X}_t^{(b),i} + K_L \left(\mathbf{Y}(k) - \mathbb{H}\mathbf{X}_t^{(b),i} \right) \tag{11}$$

with:

$$K_L = L_t^{(b)} \mathbb{H}^T \left(\mathbb{H} L_t^{(b)} \mathbb{H}^T + R_t \right)^{-1} \tag{12}$$

The values of the matrix ρ are chosen for suppressing the spurious covariances that link pixels, which should be uncorrelated otherwise. These spurious covariances appear when sampling the error covariance matrix and their values decrease if making use of large size ensembles (but this increases the computing requirements).

Let $\mathbf{p}_1$ and $\mathbf{p}_2$ denote two pixels of the domain. The value of their correlation $\rho(\mathbf{p}_1, \mathbf{p}_2)$ depends first on their distance $d_{12} = \|\mathbf{p}_1 - \mathbf{p}_2\|$ (Hamill et al., 2001). A matrix ρ_d (the subscript d is the first letter of distance) is defined with values:

$$\rho_d(\mathbf{p}_1, \mathbf{p}_2) = \left(1 + \frac{d_{12}}{a_d} \right) \exp\left(-\frac{d_{12}}{a_d} \right) \tag{13}$$

a_d is a parameter which defines the decorrelation distance: pixels apart of more than a_d are no more correlated.

The value of $\rho(\mathbf{p}_1, \mathbf{p}_2)$ should also nullify the correlation between the two pixels $\mathbf{p}_1$ and $\mathbf{p}_2$ if they belong to different structures. Let I_1 and I_2 denote their two brightness values and s_{12} be equal to $|I_1 - I_2|$. A second matrix ρ_s is defined:

$$\rho_s(\mathbf{p}_1, \mathbf{p}_2) = \left(1 + \frac{s_{12}}{a_s} \right) \exp\left(-\frac{s_{12}}{a_s} \right) \tag{14}$$

where a_s is the decorrelation parameter according to the brightness similarity between pixels.

The localization function ρ, used in Equation (10), is then defined as the product of ρ_d and ρ_s. Consequently, the correlation values between pixels that are far apart or belong to different structures are almost null, as they should have been with the analytical matrix.

The ensemble Kalman filter that is associated to the localization function ρ is named Explicit Structures Localization method, or ESL. This filter allows estimating both a dense motion field (averaging the members) and its uncertainty, and it relies on the image structures through the matrix ρ. Results obtained with that filter will be discussed in Section 5 for a direct comparison with those of the method defined in the next section.

4 Domain Decomposition

The second approach for limiting the spurious covariances, created by the approximation of $B_t^{(b)}$ through the sampling by an ensemble, is based on a decomposition of the domain Ω. The decomposition domain technique is, for instance, described in (Nerger et al., 2006), (Hunt et al., 2007) and (Janjić et al., 2011). The innovation of the paper concerns the use of the domain decomposition approach for defining a segmentation of the image acquisitions. A subdomain is corresponding to a region of the image. A dense motion field is then estimated on each subdomain with a separate ensemble Kalman filter and the results are merged on the whole image domain. As the filters are defined independently on each subdomain, the approach imposes that the only non null covariances are those between pixels of the same region (which should correspond to the displayed structures). The spurious covariances are consequently eliminated without any numerical process.

We first define the new notations that are required for describing the method. For simplifying the equations, the subscript t is suppressed in the following. The discrete image domain, which includes N_Ω pixels, is decomposed in n_D subdomains, denoted D_j for $j = 1, \ldots, \mathrm{n}_D$. The number of pixels in D_j is written N_{D_j}. We assume no overlapping between these subdomains. Con-

sequently:

$$\mathrm{N}_D = \sum_{j=1}^{\mathrm{n}_D} \mathrm{N}_{Dj} = \mathrm{N}_\Omega \tag{15}$$

Let us introduce the restriction operator $\mathcal{R}_{D_j}$: it restricts any function defined on the image domain as a function defined on the subdomain D_j. This allows defining one EnKF on each subdomain without additional notation. The merging of the results requires to define a fusion operator $\mathcal{R}_D$, which transforms any function defined on the whole acquisition domain as a vector composed of the restrictions onto all subdomains. Applying $\mathcal{R}_D$ on any function f, defined on the image domain leads to the vector f_D:

$$\mathcal{R}_D f = \begin{bmatrix} \mathcal{R}_{D_1} f \\ \vdots \\ \mathcal{R}_{D_j} f \end{bmatrix} = \begin{bmatrix} f_{D_1} \\ \vdots \\ f_{D_j} \end{bmatrix} \tag{16}$$

In the same way, the error covariance matrices $B^{(b)}$ and R are respectively replaced by the merging, denoted $B_D^{(b)}$ and R_D, of their restrictions on the subdomains indexed by j, denoted $B_{D_j}^{(b)}$ and R_{D_j}:

$$B_D^{(b)} = \mathcal{R}_D B^{(b)} \mathcal{R}_D^T \tag{17}$$

$$R_D = \mathcal{R}_D R \mathcal{R}_D^T \tag{18}$$

An additional domain decomposition E is designed for the observations. The subdomains of E are written E_j, with size N_{Ej}. We make use of the same subscript j as each observation subdomain E_j is associated to one estimation subdomain D_j. The number of observation subdomains is consequently the same than the number n_D of estimation subdomains. In order to allow a continuous (smooth) estimation of motion, the subdomains E_j usually overlap, so that the analysis computed on one pixel $\mathbf{p}$ of D_j uses observations from all its neighboring pixels, even if $\mathbf{p}$ is located at the boundary of the subdomain D_j. E_j is defined by extending the subdomain D_j with a given number of pixels in every direction. The total number of overlapping pixels N_{Ov} is equal to:

$$\mathrm{N}_{\mathrm{Ov}} = \mathrm{N}_E - \mathrm{N}_\Omega \tag{19}$$

with $\mathrm{N}_E = \sum_{j=1}^{\mathrm{n}_D} \mathrm{N}_{Ej}$. Let $\mathcal{R}_E$ denote the merging of the restrictions associated to the observation domain decomposition E, which is defined similarly to $\mathcal{R}_D$.

In the following paragraph, we illustrate these complex notations and we define the quantities $B_D^{(b)}$ and $B_E^{(b)}$ on a simple example: the image domain is decomposed into four subdomains, displayed on Figure 1.

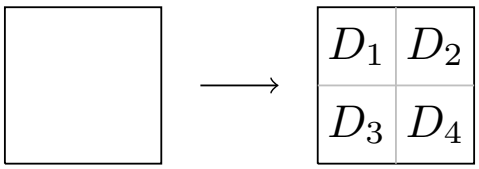

Figure 1: Decomposition of Ω into four subdomains.

The background error covariance matrix $B^{(b)}$ is first rewritten for expressing the covariances between pixels belonging to the subdomains D_i and D_j:

$$B^{(b)} = \left[B_{D_{i,j}}^{(b)} \right] \tag{20}$$

Applying the merging of restriction operators, as explained in Equation(17), gives:

$$B_D^{(b)} = \begin{bmatrix} B_{D_{1,1}}^{(b)} & 0 & 0 & 0 \\ 0 & B_{D_{2,2}}^{(b)} & 0 & 0 \\ 0 & 0 & B_{D_{3,3}}^{(b)} & 0 \\ 0 & 0 & 0 & B_{D_{4,4}}^{(b)} \end{bmatrix} \tag{21}$$

The same applies for the observation domain decomposition by replacing D by E in Equation (21).

Coming back to the general case, an ensemble Kalman filter is designed on each subdomain D_j and makes use of the values of pixels belonging to E_j. The filter is based on the following equations:

$$\mathbf{X}^{(a),i} = \mathbf{X}^{(b),i} + \mathcal{R}_D K_E \mathcal{R}_E [\mathbf{Y}(k) - \mathbb{H}\mathbf{X}^{(b),i}] \tag{22}$$

$$K_E = B_E^{(b)} \mathbb{H}^T \left(\mathbb{H} B_E^{(b)} \mathbb{H}^T + \mathcal{R}_E R \mathcal{R}_E^T \right)^{-1} \tag{23}$$

Having defined the equations of the Domain Decomposition method, we would like to understand the links between this method and the Explicit Structures Localization method, ESL.

Going back to the illustration of Fig 1 and assuming no intersection between the observation subdomains, we define the observation subdomains by $E_j = D_j$ for each value j. We define the localization function ρ:

$$\rho(\mathbf{p}_1, \mathbf{p}_2) = \begin{cases} 1 & \text{if } \exists j \text{ such as } \mathbf{p}_1 \text{ and } \mathbf{p}_2 \in D_j \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

The analysis computed by ESL is written for each member i:

$$\mathbf{X}^{(a),i} = \mathbf{X}^{(b),i} + K_L [\mathbf{Y}(k) - \mathbb{H}\mathbf{X}^{(b),i}] \tag{25}$$

$$K_L = L^{(b)}\mathbb{H}^T\left(\mathbb{H}L^{(b)}\mathbb{H}^T + R\right)^{-1} \quad (26)$$

with

$$L^{(b)} = \rho \circ B^{(b)} \quad (27)$$

It must be compared with the analysis of DD computed from Equations (22) and (23). As $E = D$, these two last equations are simplified:

$$\mathbf{X}^{(a),i} = \mathbf{X}^{(b),i} + \mathcal{R}_D K_D \mathcal{R}_D[\mathbf{Y}(k) - \mathbb{H}\mathbf{X}^{(b),i}] \quad (28)$$

$$K_D = B_D^{(b)}\mathbb{H}^T\left(\mathbb{H}B_D^{(b)}\mathbb{H}^T + \mathcal{R}_D R \mathcal{R}_D^T\right)^{-1} \quad (29)$$

If R is additionally supposed diagonal, Equation (29) is rewritten as:

$$K_D = B_D^{(b)}\mathbb{H}^T\left(\mathbb{H}B_D^{(b)}\mathbb{H}^T + R\right)^{-1} \quad (30)$$

The multiplication, in Equation (27), of $B^{(b)}$ by the function ρ, defined in Equation (24), results in a block diagonal matrix. Each block is associated to a subdomain D_j. $L^{(b)}$ is then equal to $B_D^{(b)}$ of Equation (21). Therefore, the two Kalman gains, K_D from Equation (30) and K_L from Equation (26), are equal and the two analysis values, defined by Equations (28) and (25), are the same. DD and ESL are equivalent in this simple setting.

The Domain Decomposition approach shows however major computational advantages compared to the Explicit Structures Localization. The Reader should first notice that $B_D^{(b)}$ is a block diagonal matrix and we remind that R is taken as diagonal in most applications. Therefore, the matrix to be inverted when computing the Kalman gain in Equation (23) is block diagonal. Its inverse is composed of the inverses of each block, which may be computed by independent processors in a parallel implementation. The computational cost of the matrix inversion, required when computing the Explicit Structures Localization method, is equal to $\mathrm{N}_\Omega{}^3$, while the Domain Decomposition only requires $(\sum_{j=1}^{\mathrm{n}_D} \mathrm{N}_{D_j}{}^3)$ computations, with the property:

$$\mathrm{N}_\Omega{}^3 = \left(\sum_{j=1}^{\mathrm{n}_D} \mathrm{N}_{Dj}\right)^3 \gg (\sum_{j=1}^{\mathrm{n}_D} \mathrm{N}_{D_j}{}^3) \quad (31)$$

The overlapping of the observation subdomains is mandatory for avoiding discontinuities in the result. Otherwise, two pixels $\mathbf{p}_1$ and $\mathbf{p}_2$, located on both sides of the boundary between neighboring subdomains, would be totally uncorrelated. This is illustrated by Figure 2, which is the result of motion estimation obtained without overlapping the observation subdomains. The zoom displayed on the right image illustrates the discontinuity at the boundary between two subdomains.

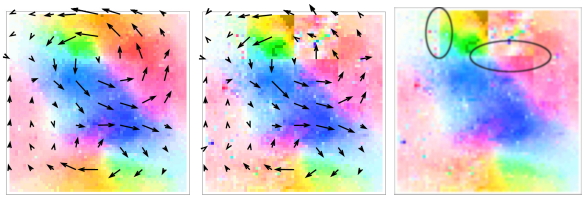

Figure 2: Left to right: Motion background $\mathbf{w}^{(b)}$; Motion estimation $\mathbf{w}^{(a)}$; Highlight on discontinuities between subdomains.

If the observation subdomains are overlapping, the Kalman gain, in Equation (23), is computed for each subdomain E_j. But the value of the estimation $\mathbf{X}^{(a),i}$ is only updated for pixels belonging to the estimation subdomain D_j. Two neighboring pixels $\mathbf{p}_1$ and $\mathbf{p}_2$, located on both sides of the two neighboring domains D_{j_1} and D_{j_2}, are updated by computations performed, respectively, on the two observation domains E_{j_1} and E_{j_2}. If E_{j_1} and E_{j_2} overlap, $\mathbf{p}_2$ belongs to E_{j_1}. When the Kalman gain is computed for pixel $\mathbf{p}_1$, it includes the covariance between $\mathbf{p}_1$ and pixels of E_{j_1}, and consequently pixel $\mathbf{p}_2$. This makes the result smoother and more significant at the boundary between the subdomains.

Unfortunately, the overlapping of observation subdomains also results in additional computations, as the Kalman gain is computed several times for the pixels belonging to the intersection between subdomains. Let $\mathrm{N}_{D\mathrm{Max}}$ denote the number of pixels of the largest subdomain D_{Max} and $\mathrm{N}_{E\mathrm{Max}}$ the number of pixels in the corresponding observation subdomain E_{Max}. The additional computation cost for each subdomain is bounded by the number of pixels included in the region $E_{\mathrm{Max}} - D_{\mathrm{Max}}$, which is supposed to be low compared to $\mathrm{N}_{D\mathrm{Max}}$. As n_D processors compute simultaneously the Kalman gain in the n_D observation subdomains, the total cost of computing the Kalman gain is a function of $\mathrm{N}_{E\mathrm{Max}}^3$, which is approximately equal to $\mathrm{N}_{D\mathrm{Max}}^3$, according to the previous assumption. This value has to be compared with the complexity $O(\mathrm{N}_\Omega{}^3)$ of the Explicit Structures Localization, with $\mathrm{N}_\Omega \gg \mathrm{N}_{D\mathrm{Max}}$. The Domain Decomposition therefore becomes an affordable approach for estimating motion with EnKF in an operational setting, as the size of the largest region may be defined according to the computational constraints.

The Domain Decomposition is usually applied, as described by Hunt *et al.* (Hunt et al.,

2007), according to regular grids. The subdomains are rectangles, whose dimensions correspond to an empirical estimation of the decorrelation value. However, the decomposition in subdomains should rely on the structures displayed on images, so that the covariances become negligible for pixels belonging to different structures. The decomposition proposed in our paper relies on a split-and-merge approach, which is a segmentation technique that has been widely used for image processing (the Reader may refer to the foundational paper of Horowitz and Pavlidis (Horowitz and Pavlidis, 1976)). The split-and-merge segmentation is based on a quadtree partition of the image. Starting from the whole image, the splitting is iterated as long as each new region is heterogeneous. A merging phase is then applied, in which regions with similar properties are combined. The merging phase aims to reduce the number n_D of subdomains. An extended observation subdomain E_i is then associated to each subdomain D_i in order to impose the smoothness property, as it has been previously justified. Having minimized the number n_D of subdomains consequently minimizes the number N_{Ov} of overlapping pixels in the observations subdomains. The parameters involved in the split-and-merge method are related to the distance and the similarity between pixels. The first parameter defines the maximal size of an analysis subdomain D_j. If we want to compare the domain decomposition method to the ESL technique, this maximal size is chosen so that the distance between two pixels belonging to D_j is lower than the decorrelation value used with the ESL approach. The second parameter defines the notion of homogeneous region. In other words, it defines which properties should be verified by the pixels of an analysis subdomain D_j, so that D_j is considered as homogeneous. The classical choice is to consider that the standard deviation σ_j of the gray level values within the subdomain D_j should be smaller than a given threshold. In order to compare with the ESL method, the value of σ_j is chosen to be equal to the brightness decorrelation value of ESL.

5 Description of the experiments

This section displays and discusses results obtained with both the ESL and DD approaches, but we remind that only DD is also applied in an operational setting as the computation is done in parallel for each subdomain. The two methods are set up with the same criteria on the size of regions and on the similarity of gray level values, in order to get an objective comparison of results with the same modeling of structures.

A last component has to be defined which is the model of dynamics $\mathbb{M}$ used in Equation (4). In our experiments, this model assume the Lagrangian constancy of velocity on each pixel trajectory and the transport of the image brightness by velocity:

$$\frac{\partial \mathbf{w}}{\partial t}(\mathbf{x},t) + (\mathbf{w}.\nabla)\mathbf{w}(\mathbf{x},t) = 0 \qquad (32)$$

$$\frac{\partial I}{\partial t}(\mathbf{x},t) + \mathbf{w}.\nabla I(\mathbf{x},t) = 0 \qquad (33)$$

The next subsections discuss first the results on synthetic data and then demonstrate its properties on satellite acquisitions.

5.1 Synthetic experiment

The first experiment relies on a sequence of 9 synthetic images, which have characteristics similar to satellite acquisitions. These images are obtained by integrating in time, with the model $\mathbb{M}$ of Equations (32) and (33), the motion field displayed on top of Figure 4 and the satellite image on top of Figure 3. This process allows, not only to obtain the sequence of image observations used for testing motion estimation methods, but also to compute the ground truth on motion. The bottom parts of the two previous Figures display the last image of the sequence and the ground truth on motion at the corresponding time.

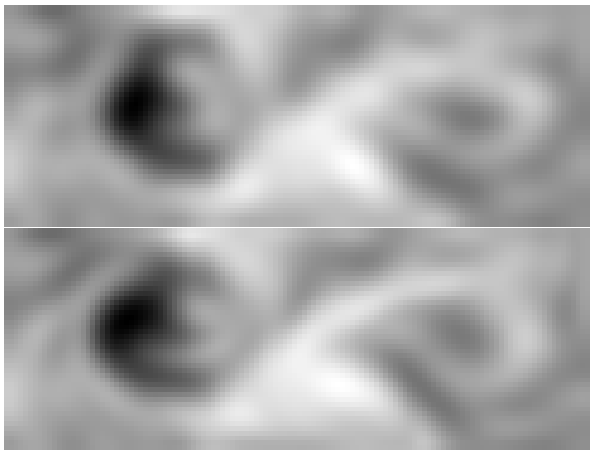

Figure 3: First and last image of the synthetic sequence.

In the following, we discuss the results obtained with our two methods, ESL and DD, and

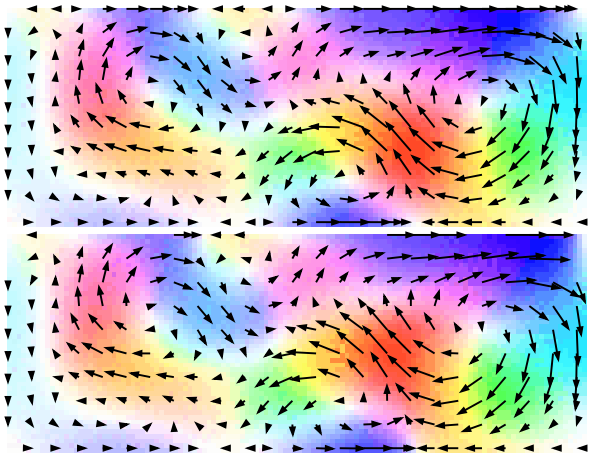

Figure 4: Ground truth on motion.

compare with those of the HS-brightness algorithm (Sun et al., 2010), which gets the best performances among all methods, which have been tested on our experimental data.

Figure 5 provide the results that are computed at the beginning of the sequence, while Figure 6 illustrates the results obtained at the last time. It should be noted that, in the first case, HS-brightness estimates motion from images 1 and 2, while in the second case this is between images 8 and 9.

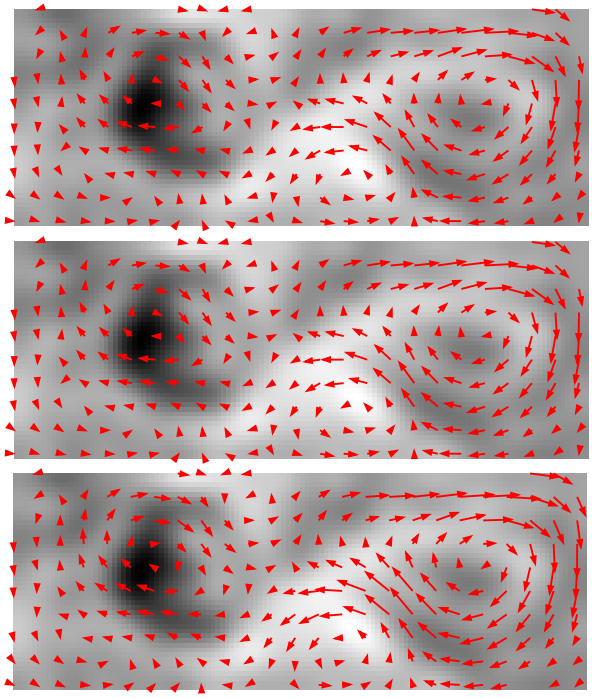

Figure 5: Result of ESL (top), DD (bottom) and HS-brightness (bottom) on the first image.

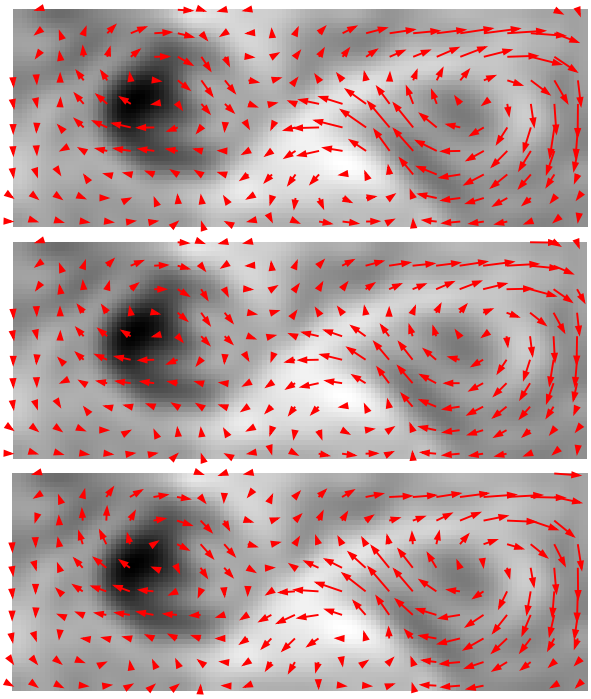

Figure 6: Result of ESL (top), DD (middle) and HS-brightness (bottom) on the last image.

Figure 7 displays the domain decomposition that is used for the DD approach, which is regular in that experiment.

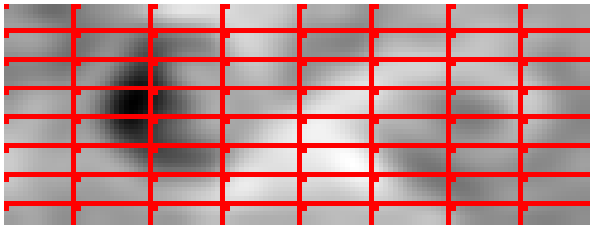

Figure 7: Domain decomposition.

The motion estimation obtained with the three methods are visually equivalent and close from the ground truth displayed on Figure 4. For an improved understanding of the methods, we computed statistics on the errors of the estimations with regard to the ground truth. At initial time, the angular errors of the three methods ESL, DD and HS-brightness are approximately the same and around 9 degrees. However, at the last time, the angular errors of the three methods are different and respectively of 9.0, 9.6 and 10.33 degrees for ESL, DD and HS-brightness. HS-brightness is one of the initial member of the ensemble and it has demonstrated to be the best member. However, the other members are con-

tributing during time integration to that member as we previously discussed. At the last time, ESL and DD perform better than their best member. That means that all methods are contributing in the results. Apart of the result quality, the main advantage of DD compared to ESL concerns the computational time, which is divided by the number of subdomains, which is 64 for this illustration.

One additional output of the ESL and DD methods, compared to HS-brightness, is the uncertainty of the motion estimation, as discussed in Sections 3 and 4. We first design an image of the uncertainty on the orientation of motion at each time t. Let $\mathbf{w}_e^i(\mathbf{x},t)$ be the value of the i^{th} member, at location $\mathbf{x}$ and time t, and $\mathbf{w}_e(\mathbf{x},t)$ be the estimation of the ensemble Kalman filter, obtained as the average of all members values. We define $U(\mathbf{x},t) = \overline{\widehat{\mathbf{w}_e^i(\mathbf{x},t),\mathbf{w}_e}(\mathbf{x},t)}$. This is the average of the angular difference between each member i and the estimation. This characterizes the angular spread of the ensemble at pixel $\mathbf{x}$ and time t. We also compute the map of angular errors, $E(\mathbf{x},t) = \widehat{\mathbf{w}_{GT}(\mathbf{x},t),\mathbf{w}_e}(\mathbf{x},t)$, between the ground truth $\mathbf{w}_{GT}(\mathbf{x},t)$ and the estimation $\mathbf{w}_e(\mathbf{x},t)$. This process is illustrated on Figures 8 and 9 where the two images $U(\mathbf{x},t)$ and $E(\mathbf{x},t)$, computed for the method DD at the first and last time of the sequence, are displayed. Results are similar with the method ESL. It is visible that the uncertainty measured by the ensemble is linked to the error of the estimation. Further analysis is part of the perspectives of this research.

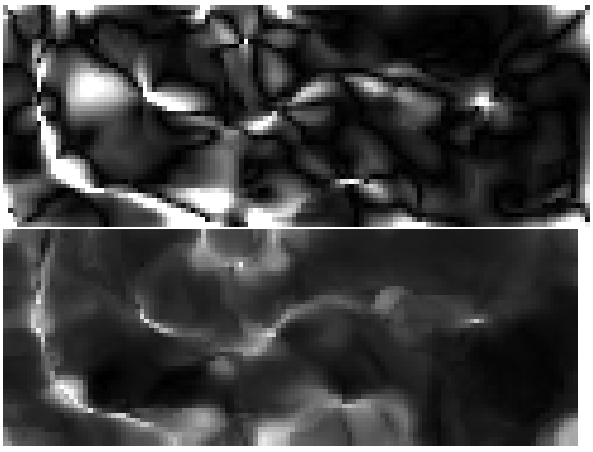

Figure 8: Angular error map of DD, E_{DD}, (top) and uncertainty of angular error of DD, U_{DD}, on the first image.

Having analyzed the results on synthetic data, we will demonstrate the potentiality of ESL and DD on satellite acquisitions, in the next subsection.

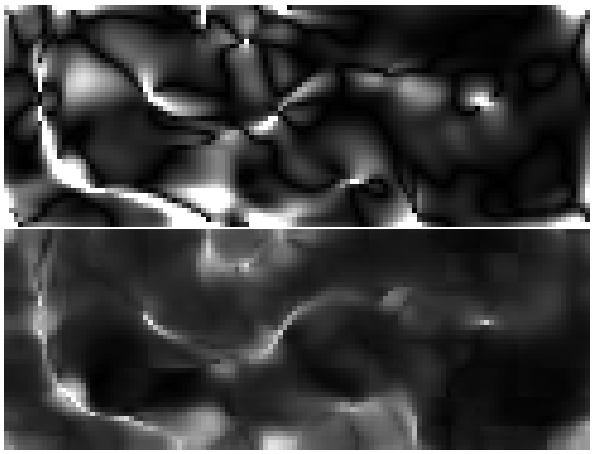

Figure 9: Angular error $E(\mathbf{x})$ and uncertainty on the orientation $U(\mathbf{x})$, with DD at last time.

5.2 Satellite data

Experiments were further conducted with meteorological satellite sequences, such as the one displayed on Figure 10 (frames 1, 4, 8, 11, 15, 18).

The domain decomposition obtained with the split-and-merge algorithm is given on Figure 11. This gives a rough segmentation on the image, as the boundaries of the clouds structures are not accurately defined.

Motion results obtained with the ESL, DD and HS-brightness methods are visualized on Figure 12 for the initial time and Figure 13 for the last time. Motion results of ESL and DD are equivalent, but the computational time required by DD is 55 less than the one of ESL, since the split-and-merge divides the whole domain in 55 subdomains. Results of HS-brightness show that this approach is less suited on the real satellite data. As HS-brightness is the best of the 50 members used in the two methods, this also demonstrates again that EnKF makes use of all its members and that all methods are contributing to the final estimation.

Figure 14 shows the uncertainty map, previously named image $U(\mathbf{x})$, computed for the DD method at the first time.

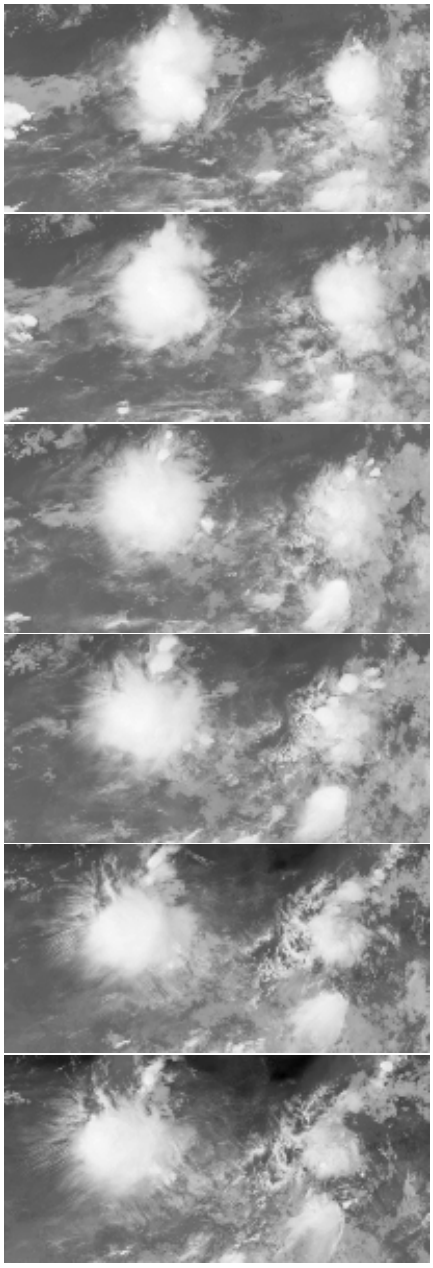

Figure 10: Six satellite acquisitions.

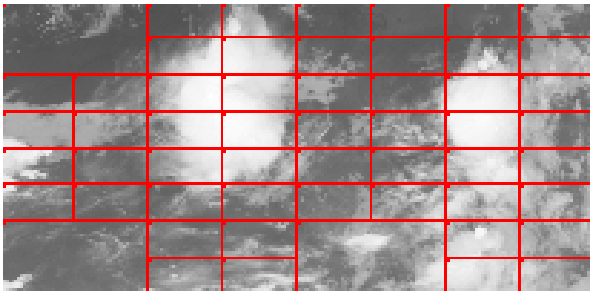

Figure 11: Domain decomposition performed by DD on the first observation.

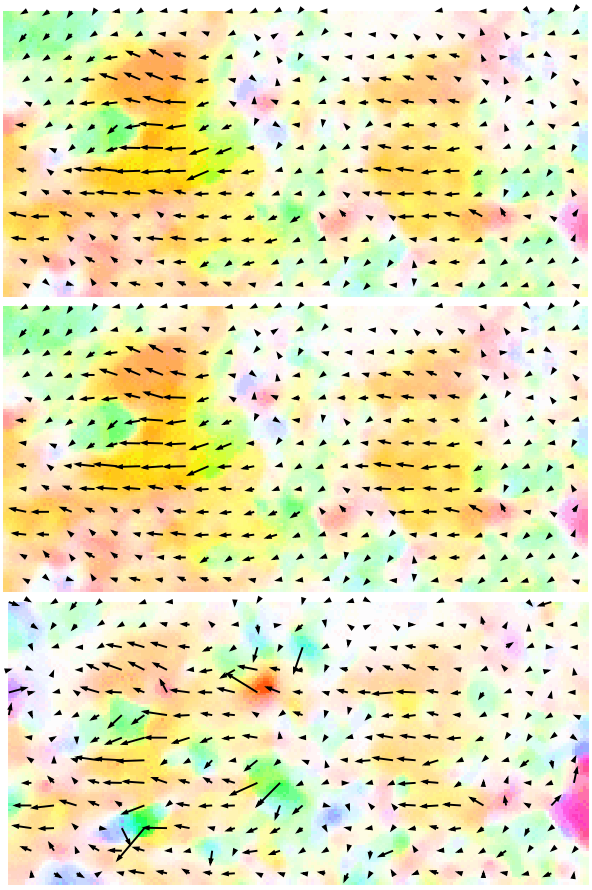

Figure 12: Comparison of ESL (top), DD (middle) and HS-brightness (bottom) at initial time.

6 Conclusions

This paper describes the design of an ensemble Kalman filter that estimates motion on fluid flows images and provides a measure on the uncertainty of the result. For suppressing the spurious covariances, coming from the sampling of the PDF by an ensemble, two methods are compared which rely on the structures displayed on the data. On the one hand, the Explicit Structures Localization method, ESL, which was defined in the past by the authors, makes use of a localization function for keeping correlated the only pixels, which are simultaneously close (distance property) and having similar gray level values (brightness property). On the other hand, the innovation of the paper concerns a Domain Decomposition approach, named DD, which is based on a rough segmentation of the image and allows processing each subdomain in parallel. The two methods have been proven to be equivalent on synthetic and satellite data, but only the Domain

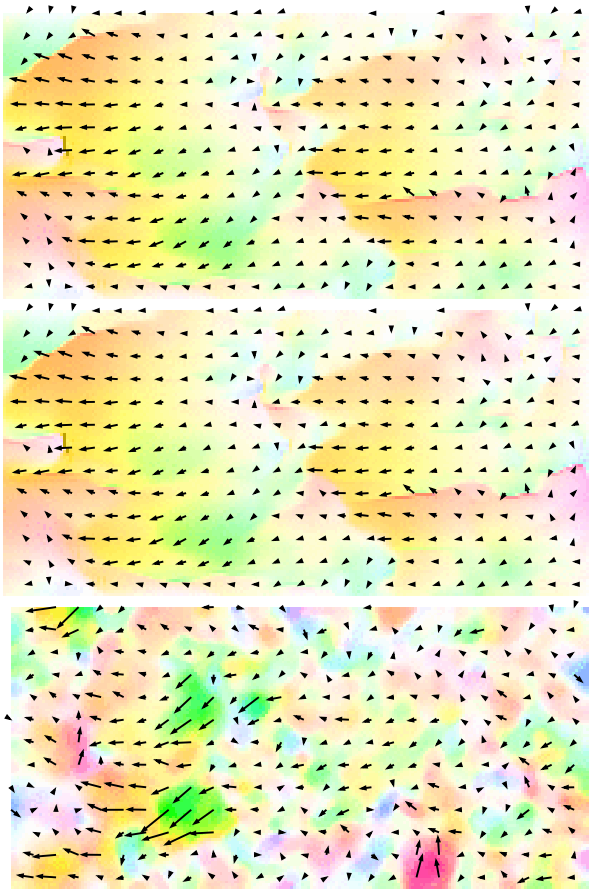

Figure 13: Comparison of ESL (top), DD (middle) and HS-brightness (bottom) at last time.

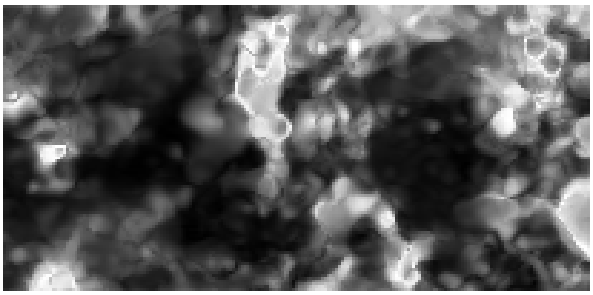

Figure 14: Uncertainty computed by DD at first time.

Decomposition approach is applicable in an operational setting, due to its reduced computational requirements.

Various perspectives are considered for this research. First, an improved space-time decomposition of the domain will be implemented as this is the key point for both decreasing the number of subdomains (and decreasing the computational cost), while correctly characterizing the structures. On a methodological point of view, the research will consider approaches such as LETKF (Local Ensemble Transform Kalman Filter, see (Miyoshi et al., 2007)) for improving the estimation of the uncertainty computed from the ensemble. Last, the use of the uncertainty for operational applications such as rain quantity forecast will be further assessed.

REFERENCES

Anderson, J. L. (2007). An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus Series A : Dynamic meteorology and oceanography*, 59:210–224.

Anderson, J. L. (2001). An Ensemble Adjustment Kalman Filter for Data Assimilation. *Monthly Weather Review*, 129(12):2884–2903.

Baker, S., Scharstein, D., Lewis, J. P., Roth, S., Black, M. J., and Szeliski, R. (2011). A database and evaluation methodology for optical flow. *International Journal on Computer Vision*, 92(1):1–31.

Elad, M. and Feuer, A. (1998). Recursive optical flow estimation – adaptive filtering approach. *Journal of Visual Communication and Image Representation*, 9(2):199–138.

Evensen, G. (2003). The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, 53:343–367.

Franke, U. and Rabe, C. (2005). Kalman filter based depth from motion with fast convergence. In *Intelligent Vehicles Symposium*.

Franke, U., Rabe, C., Badino, H., and Gehrig, S. (2005). 6d-vision: Fusion of stereo and motion for robust environment perception. In *DAGM Symposium*.

Hamill, T. M., Whitaker, J. S., and Snyder, C. (2001). Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Monthly Weather Review*, 129:2776–2790.

Horn, B. and Schunk, B. (1981). Determining optical flow. *Artificial Intelligence*, 17:185–203.

Horowitz, S. L. and Pavlidis, T. (1976). Picture segmentation by a tree traversal algorithm. *Journal ACM*, 23(2):368–388.

Houtekamer, P. L. and Mitchell, H. L. (1998). A sequential ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 129:123–137.

Hunt, B. R., Kostelich, E. J., and Szunyogh, I. (2007). Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena*, 230(1):112–126.

Janjić, T., Nerger, L., Albertella, A., Schröter, J., and Skachko, S. (2011). On domain localization in ensemble-based Kalman filter algorithms. *Monthly Weather Review*, 139(7):2046–2060.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45.

Lepoittevin, Y., Herlin, I., and Béréziat, D. (2015). An image-based ensemble kalman filter for motion estimation. In *International Conference on Computer Vision Theory and Applications*.

Lowe, D. (1999). Object recognition from local scale-invariant features. In *International Conference on Computer Vision*.

Miyoshi, T., Yamane, S., and Enomoto, T. (2007). Localizing the error covariance by physical distances within a local ensemble transform Kalman filter (LETKF). *Scientific Online Letters on the Atmosphere*, 3:8992.

Nerger, L., Danilov, S., Hiller, W., and Schröter, J. (2006). Using sea-level data to constrain a finite-element primitive-equation ocean model with a local SEIK filter. *Ocean Dynamics*, 56(5-6):634–649.

Oke, P. R., Sakov, P., and Corney, S. P. (2007). Impacts of localisation in the EnKF and EnOI: experiments with a small model. *Ocean Dynamics*, 57(1):32–45.

Rabe, C., Mller, T., Wedel, A., and Franke, U. (2010). Dense, robust, and accurate motion field estimation from stereo image sequences in real-time. In *European Conference on Computer Vision*.

Sun, D., Roth, S., and Black, M. (2010). Secrets of optical flow estimation and their principles. In *European Conference on Computer Vision*, pages 2432–2439.

Tomasi, C. and Kanade, T. (1991). Detection and tracking of point features. Technical Report CMU-CS-91-132, Carngie Mellon University.