Dealing with convergence problems when accounting for correlated observation errors in image assimilation

Vincent Chabot[1]
Maëlle Nodet[2], Arthur Vidard[3]

[1]Météo-France, RTRA-STAE
[2]Université de Grenoble
[3]INRIA Grenoble - Rhône-Alpes

June 2, 2015

Motivation

- Error in dense field, such as satellite images, are correlated in space.
- Model resolutions are increasing. Need to extract finer structure from observation.
- Observation error covariance matrices are large and block diagonal.

Hypothesis (in this talk):

- The true **R** matrix is known.
- The observations are only correlated in space.

Questions:

- How to use this information in a 4D-Var?
- What kind of issue could arise? Why?

Outline

1. **Modeling R through a change of variable**
2. **Experiments with an isotropic noise**
3. **Convergence issue**

Technical problems regarding $\mathbf{R}$ matrix

Algorithm: 4D-Var with $\mathbf{B}^{1/2}$ preconditioning.
Problem: Need to compute $\mathbf{R}^{-1}(\mathbf{y} - \mathcal{H}\mathbf{x})$ at each iteration.

Constraints:

- $\mathbf{R}$ should be invertible,
- the product $\mathbf{R}^{-1}(\mathbf{y} - \mathcal{H}\mathbf{x})$ should not be too expensive.

For dense field, we can use methods similar to those developed for $\mathbf{B}$ matrix.

Main differences:

- $\mathbf{R}$ needs to be inverted,
- the observation space changes with time.

Representation of spatial correlation in R through a change of variable

There are different ways to represent spatial correlation (**[Fisher 2003]**, **[Stewart et al. 2013]**, **[Weaver 2014]**, ...).
In this talk, we use a diagonal matrix after a change of variable (see **[Chabot et al. 2014]**).

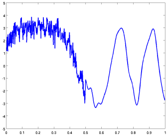
Suppose $y^o = y^t + \epsilon$ with $\epsilon \sim \mathcal{N}(0, \mathbf{R})$.
Then $\mathbf{A}y^o = \mathbf{A}y^t + \beta$ with $\beta \sim \mathcal{N}(0, \mathbf{ARA}^T)$.

Aim

Choose $\mathbf{A}$ such that $\mathbf{D_A} = \mathrm{diag}(\mathbf{ARA}^T) \simeq \mathbf{ARA}^T$.

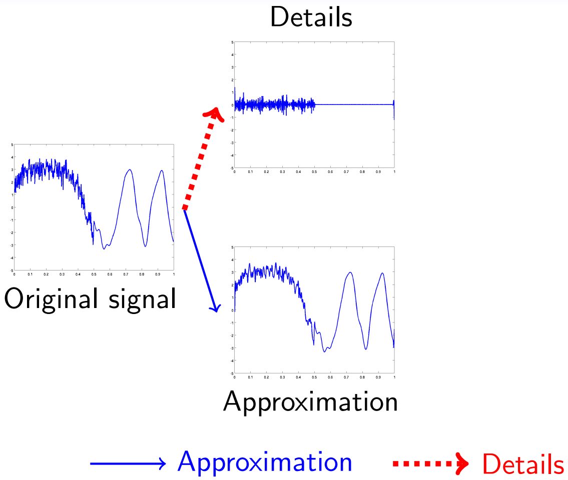
Here **A** is an orthonormal wavelet transform.

Spirit of a wavelet transform

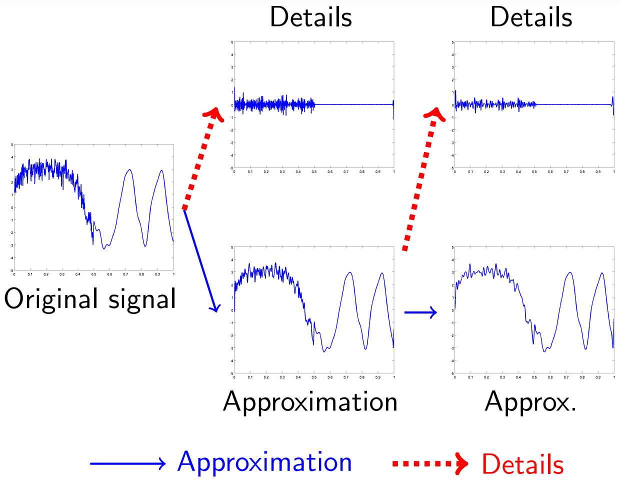


Original signal

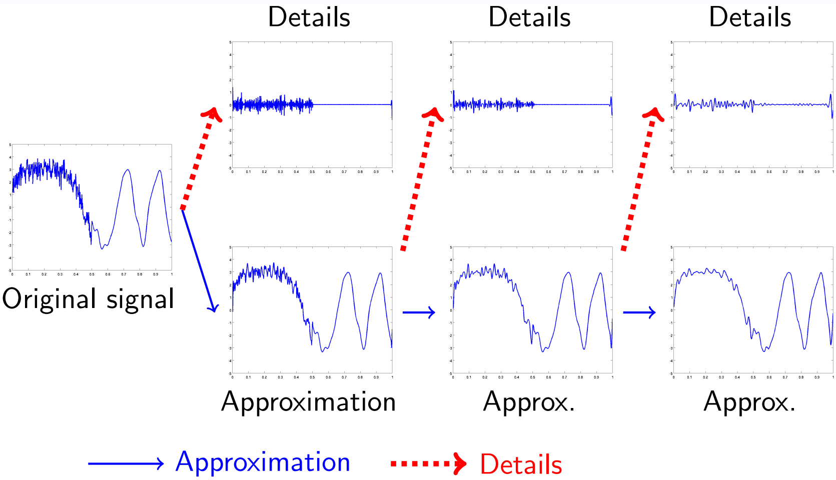
⟶ Approximation ⇢ Details


Details
Original signal
Approximation
Approximation
Details

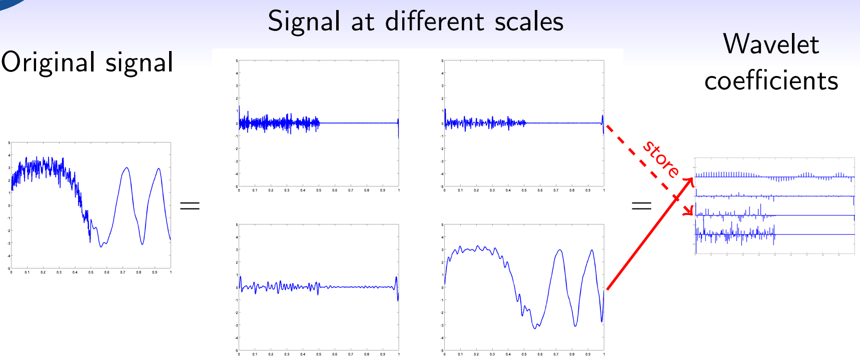
Spirit of a wavelet transform



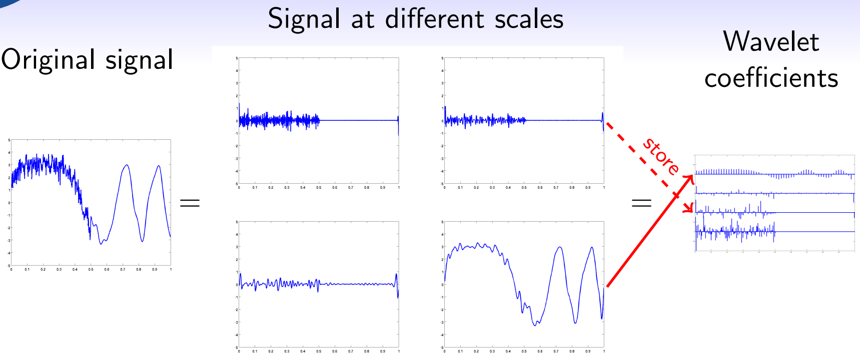
Spirit of a wavelet transform



Spirit of a wavelet transform



Spirit of a wavelet transform



Summary

Use of a "basis" where each element has some **scale**, **orientation** and **spatial localization** properties. Write the cost function as:

$$(\mathbf{y}-\mathcal{H}(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{y}-\mathcal{H}(\mathbf{x})) = (\mathbf{y}-\mathcal{H}(\mathbf{x}))^T\mathbf{A}^T\mathbf{D}_{\mathbf{A}}^{-1}\mathbf{A}(\mathbf{y}-\mathcal{H}(\mathbf{x}))$$

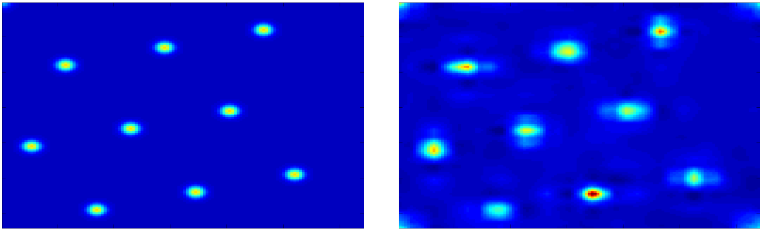
Return in pixel space

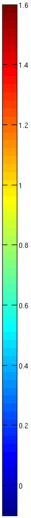
Go in wavelet space

Divide by the variance

Example of covariance matrix : isotropic and homogeneous case

True | Diagonal wavelet modelisation





Orthonormal wavelet transforms do not preserve (in general):

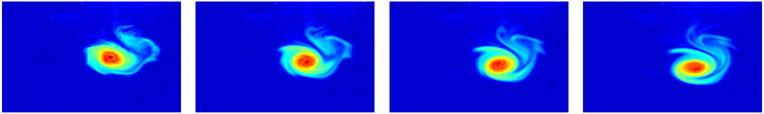
- the variance value (in pixel space),
- the spatial localization,
- the isotropy or the homogeneity of the covariance fields,

but enable to represent (at a cheap cost) some of the error correlations.

Outline

Twin experiment context

Model : Shallow-water $\Rightarrow$ quantities of interest are (u,v,h)
Observations : an image sequence of passive tracer $\Rightarrow \mathcal{H}$ is modelled by an advection–diffusion equation.



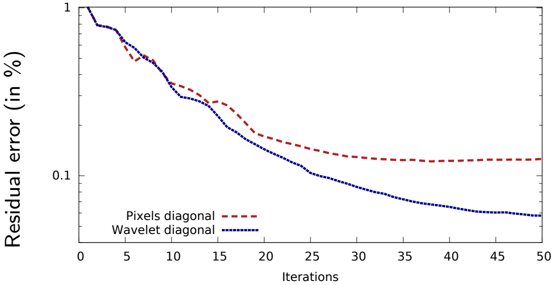
Algorithm : 4D-Var with $\mathbf{B}^{1/2}$ preconditioning.
B is modeled by diffusion operators **[see Weaver and Courtier 2001]**.
Background : $(u_0, v_0, h_0) = (0, 0, h_{mean})$

Aim

Control the velocity field via the assimilation of a noisy passive tracer sequence.

Results with homogeneous isotropic noise

Observations : $\mathbf{y^o_{ti}} = \mathbf{y^t_{ti}} + \epsilon_{\mathbf{iso}}$



- Accounting for error correlations leads to a decrease of the residual error.
- There is no convergence issue in this case.

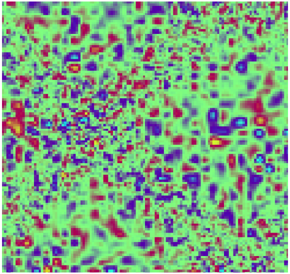
Outline

Convergence issue : best matrix representation in a wavelet space

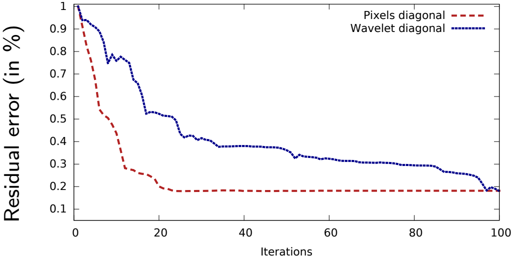
The true covariance matrix is used in the wavelet space

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

A noise realization



RMSE with respect to the minimization iterations

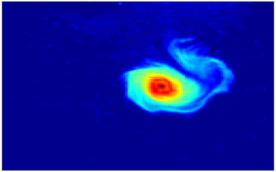


Incorporating the true covariance information leads to some convergence issue.

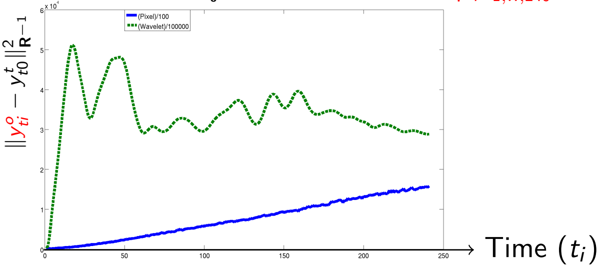
Using the multiscale aspect of the Wavelet transform

What happens when discarding information from small scales?

Full image



Distance between $\mathbf{y}^{\mathbf{t}}_{\mathbf{t_0}}$ and the observations $\left(\mathbf{y}^{\mathbf{o}}_{\mathbf{t_i}}\right)_{i=1,..,240}$



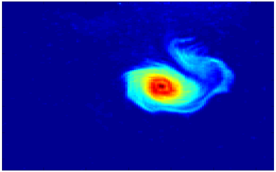
— (Pixel Diagonal)/100

— (Wavelet)/100000

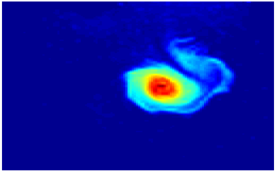
Using the multiscale aspect of the Wavelet transform

What happens when discarding information from small scales?

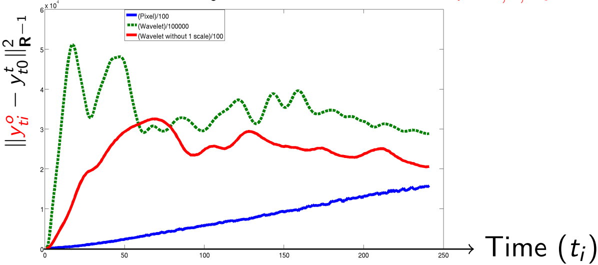
Full image



Discard 1 scale



Distance between $\mathbf{y}^{\mathbf{t}}_{\mathbf{t_0}}$ and the observations $\left(\mathbf{y}^{\mathbf{o}}_{\mathbf{t_i}}\right)_{i=1,..,240}$



— (Pixel Diagonal)/100

— (Wavelet)/100000

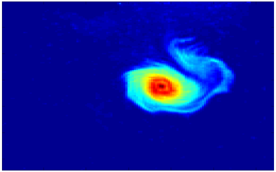
— (Wavelet without 1 scale)/100

Discarding some information enables to get a better distance notion.

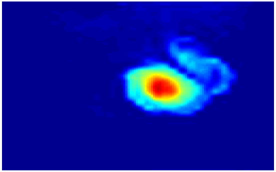
Using the multiscale aspect of the Wavelet transform

What happens when discarding information from small scales?

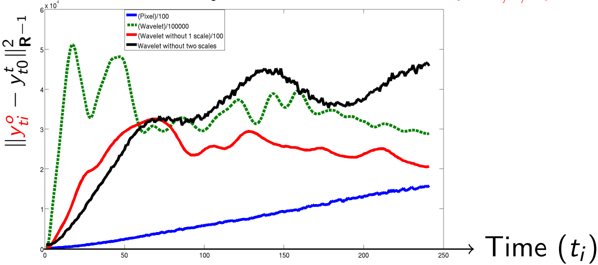
Full image



Discard 2 scales



Distance between $\mathbf{y}^{\mathbf{t}}_{\mathbf{t_0}}$ and the observations $\left(\mathbf{y}^{\mathbf{o}}_{\mathbf{t_i}}\right)_{i=1,..,240}$



— (Pixel Diagonal)/100 — Wavelet without 2 scales

— (Wavelet)/100000

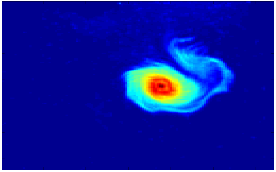
— (Wavelet without 1 scale)/100

Discarding some information enables to get a better distance notion.

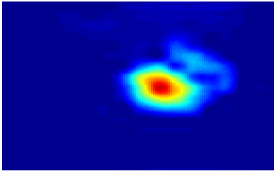
Using the multiscale aspect of the Wavelet transform

What happens when discarding information from small scales?

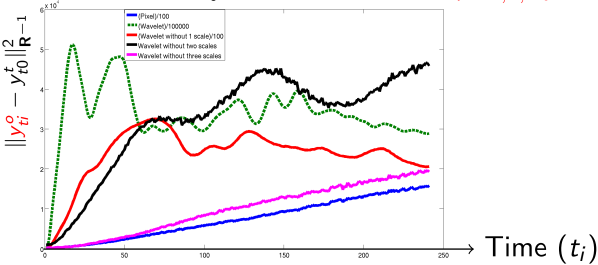
Full image



Discard 3 scales



Distance between $\mathbf{y}^{\mathbf{t}}_{\mathbf{t_0}}$ and the observations $\left(\mathbf{y}^{\mathbf{o}}_{\mathbf{t_i}}\right)_{i=1,..,240}$



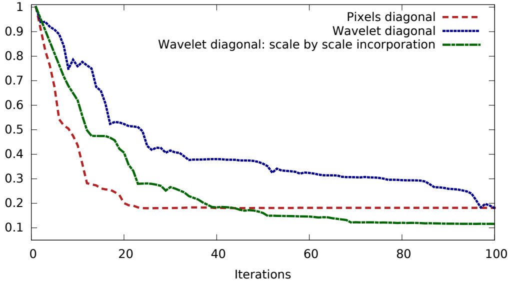
— (Pixel Diagonal)/100
— (Wavelet)/100000
— (Wavelet without 1 scale)/100
— Wavelet without 2 scales
— Wavelet without 3 scales

Discarding some information enables to get a better distance notion.

Accelerate the convergence rate

Idea

Use only coarsest information at the beginning of the minimization.
Along the minimization process, incorporate more and more information on fine scale.



It accelerates the convergence.

Conclusion and Future work

Conclusion

- It is possible to incorporate spatial error correlations through a change of variable.
- This can have some positive impact on the assimilation process.
- This can induced some convergence issues.
- It is possible to overcome this by discarding small scale information at the beginning of the assimilation process.

Future work

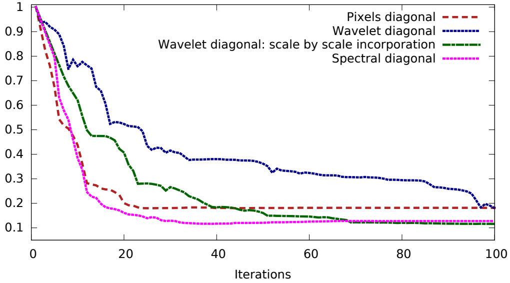
- **R** formulation in a wavelet space without full image.
- Study the impact of temporal correlation.

Questions ?

Accelerate the convergence rate

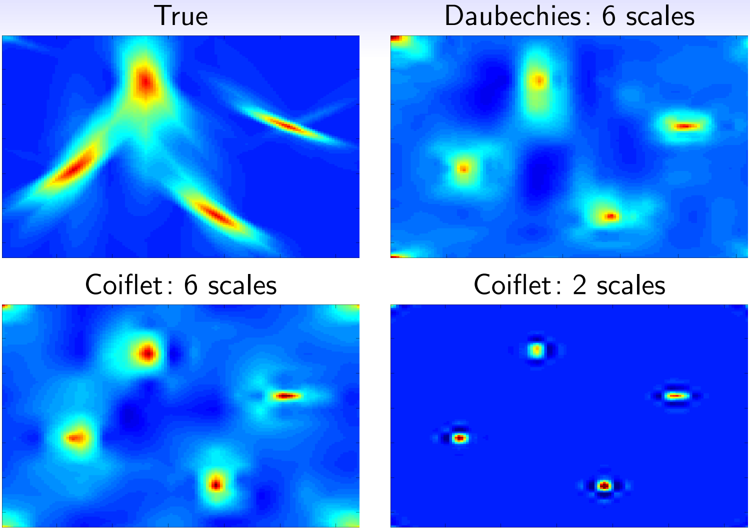
Idea

Use only coarsest information at the beginning of the minimization.
Along the minimization process, incorporate more and more information on fine scale.

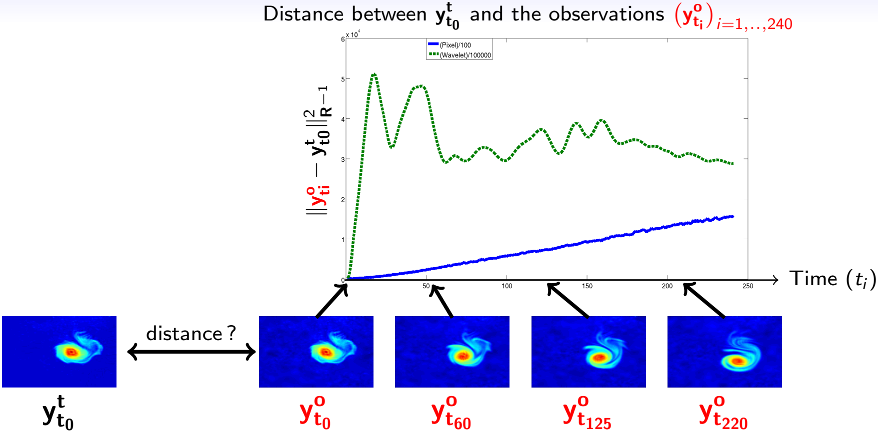


It accelerates the convergence $\cdots$ up to a certain point.

Example : inhomogeneous case



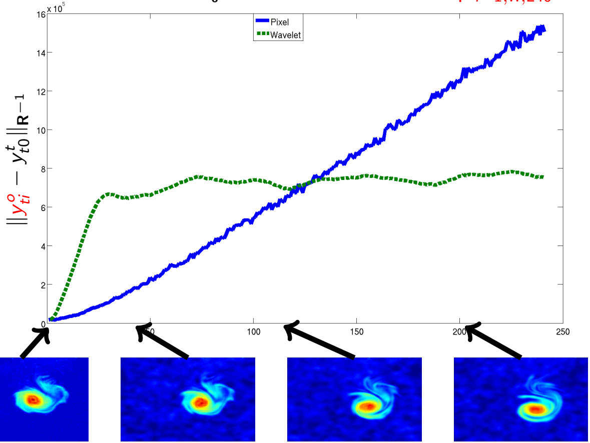
Issue with the distance



The order induced by the wavelet distance (which takes into account error correlations) is not the one expected.

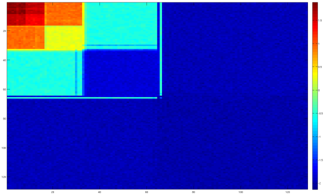
Issue with the distance : an homogeneous isotropic case

Distance between $\mathbf{y}^{\mathbf{t}}_{\mathbf{t_0}}$ and the observations $\left(\mathbf{y}^{\mathbf{o}}_{\mathbf{t_i}}\right)_{i=1,..,240}$



Variance value in Wavelet Space

Isotropic case: $log_{10}(\sigma^2)$



Inhomogeneous case: $log_{10}(\sigma^2)$