Accounting for (some) correlated observation errors in image data assimilation

Vincent Chabot, Maëlle Nodet, Arthur Vidard

LABORATOIRE JEAN KUNTZMANN
MATHÉMATIQUES APPLIQUÉES - INFORMATIQUE

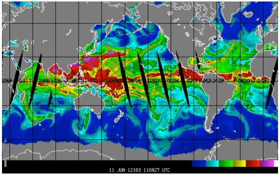
Inria informatics mathematics

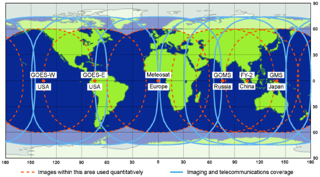
UNIVERSITÉ GRENOBLE ALPES

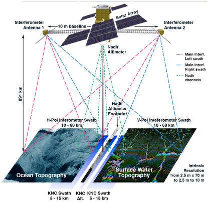
Peyresq, $25^{th} - 26^{th}$ of April 2015

Peyresq Summer School 2015

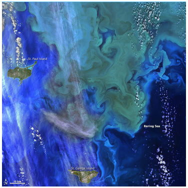
Un peu tout et n'importe quoi



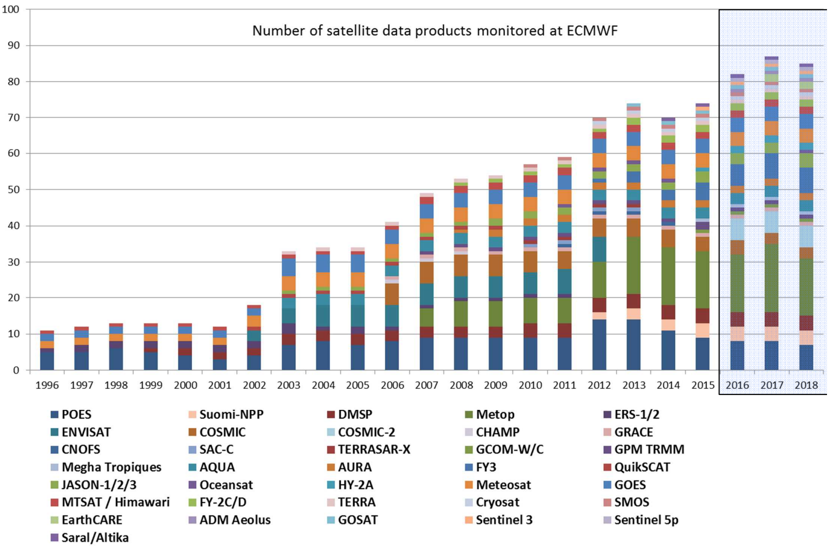




- Heterogeneous in quality, quantity, nature.
- Sparse under the surface
- Continuously increasing number.



Une quantité d'observation en constante augmentation



Currently NWP centers receive 300 million observations from 130 sources daily.

Et les corrélations spatiales (et temporelles)

- Both measurement and representativity errors
- For satellite data both are correlated in space (and time, and between channels, and ...).
  - measurement: same device, preprocessing
  - Note: representativity error are always correlated
- Number of observation is increasing, therefore **R** is getting larger.
  - Computing issue: for 3D/4DVar, we need its inverse (actually Inverse $\times$ vector)
  - Storage issue: even if sparse, its inverse is not.

Storing and inverting **R** is generally out of reach, so people try to get rid of the correlation by

- (Almost) ignoring the problem *i.e.* keeping **R** diagonal but with inflated variances
- Removing/altering part of the signal (subsampling or superobbing)
- Modelling **R** as an operator and/or change of variable.

$$\mathbf{R} = \mathbf{W}^T \mathbf{R}_{wav} \mathbf{W} \approx \mathbf{W}^T \Sigma_{wav}^2 \mathbf{W}$$

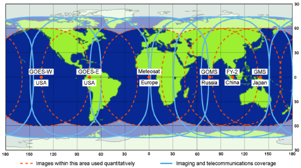
Outline

Outline

Original motivations

From geostationary satellites

Geostationnary satellites cover



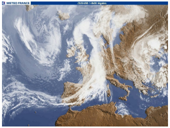
- High-resolution observations
- Quasi-global covering
- Spatially consistent
- Very expensive to acquire

Current use in NWP systems

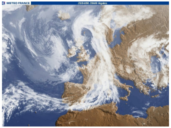
- qualitative interpretation
- as pseudo-observations (AMVs)
- as point-wise radiances

Satellite images potential

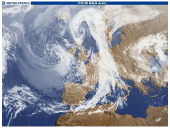
- Multiscale information on the movement (global and local).
- The images contains lagrangian tracers.
- The information is borne by the discontinuities in the field of the images (fronts, vortices).



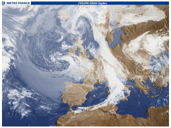
April 28, 2008, 14H00



April 28, 2008, 20h00



April 29, 2008, 02H00



April 28, 2008, 08h00

Motivation: include image sequence type information directly into the data assimilation schemes

Satellite images
From Polar-orbiting Satellites

Fancy movie about Salinity buildup

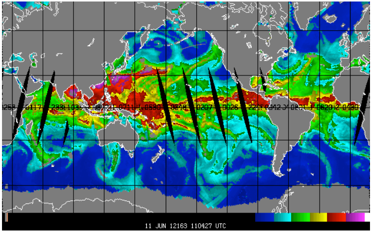
Nice movie about SMOS measures

SMOS trajectory (animations stolen from ESA website)

Image (sequence): Observation structured in space (and time)

Difficulties

Partial and composite images

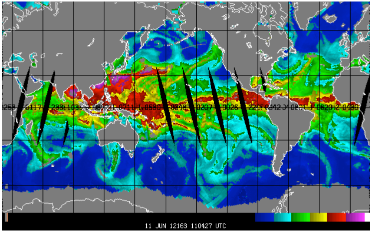


Total Precipitable Water (source CIRA)

Tigerbunny (source D. Langley)

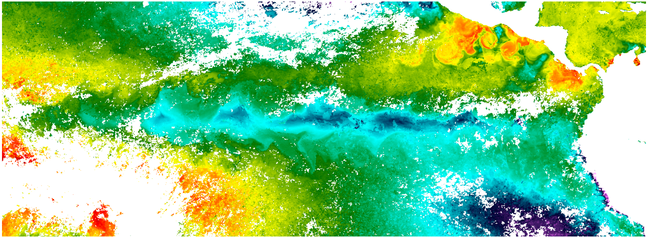
Difficulties

Partial and composite images



Total Precipitable Water (source CIRA)

Tigerbunny (source D. Langley)



SST composite sequence from metop/AVHRR Jan-Fev 2008

Difficulties

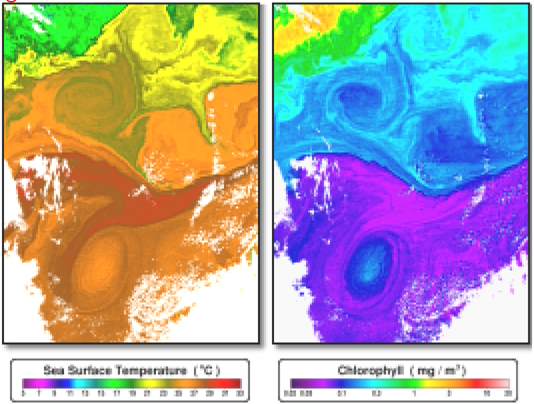
Non exhaustive list:

- The state variables of the numerical models may not be directly measured by satellites.
- The physical processes that are observed are not always taken into account in the model.
- Satellite images can be of relatively poor quality, this is mostly true for ocean surface images, which are very often partially occulted by clouds. Moreover they are most of the time composite images
- The luminance level may vary during an image sequence
- The massive amount of data makes it difficult to handle such observations.
- Images are bidimensionnal information whereas physical processes of geophysical fluids are three dimensional.
- The pertinent informations coming from an image are mainly brought by its discontinuities or high gradients. Unfortunately, numerical models have a tendency to smooth these discontinuities out.
- What can be observe is only apparent motion (beware of aliasing).

Interpretation levels of images

Elementary level and structured level

- **Elementary** level: value at each pixel
  - Large quantity of information
  - Strong dependance from acquisition conditions and measuring uncertainties

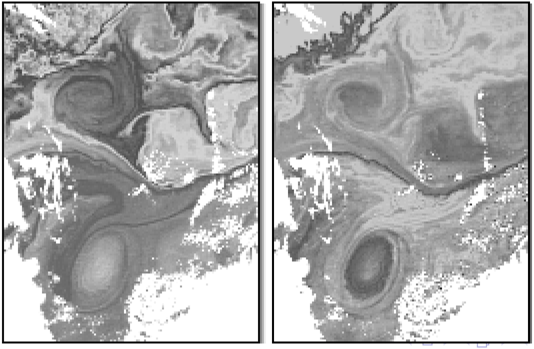


(image MODIS, source NASA)

Interpretation levels of images

Elementary level and structured level

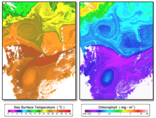
- **Elementary** level: value at each pixel
- **Structured (global)** level: spacial organization of pixels
  - Weak dependance from **acquisition conditions** and measuring **uncertainties** (but not representativeness)
  - Dominated by the global dynamics



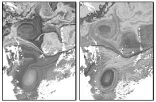
Interpretation levels of images

Elementary level and structured level

- **Elementary** level: value at each pixel
  - Large quantity of information
  - Strong dependance from acquisition conditions and measuring uncertainties



- **Structured (global)** level: spacial organization of pixels
  - Weak dependance from **acquisition conditions** and measuring **uncertainties** (but not representativeness)
  - Dominated by the global dynamics



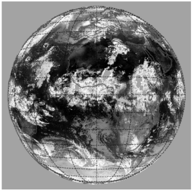
⇒ Two possible level of interpretation for the images

Data assimilation requirements and specific issues

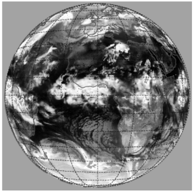
We need:

- **Observation operator**
  - Pseudo-observations (AMVs, Optical flow)
  - Use of proxies (e.g. Lyapunov vectors/exponents).
  - Contour tracking (Snakes, Levelset, bogus,...).
  - Create a synthetic image from model output.
- **Discrepancy measure**
  - L^2 norm (comparing pixels)
  - H^1 semi-norm (comparing gradients)
  - multi-scale decomposition
  - H^{-1}, W_2
- **Error statistics description**
  - Observation errors are correlated
  - The number of observation is large

Meteosat 9 IR10.8 20080525 0 UTC



RTTOV gen. Meteosat 8 IR10.8 ECMWF Fc 20080525 00 UTC:



Obs and Model

(source ECMWF)

Specific issues:

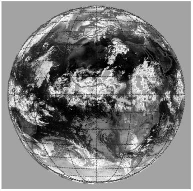
- **Occlusions and coasts** (ocean)
- **2D representation of a 3D observed process** (atmosphere)

Data assimilation requirements and specific issues

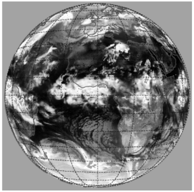
We need:

- **Observation operator**
  - Pseudo-observations (AMVs, Optical flow)
  - Use of proxies (e.g. Lyapunov vectors/exponents).
  - Contour tracking (Snakes, Levelset, bogus,. . .).
  - Create a synthetic image from model output.
- **Discrepancy measure**
  - L^2 norm (comparing pixels)
  - H^1 semi-norm (comparing gradients)
  - multi-scale decomposition
  - H^{-1}, W_2
- **Error statistics description**
  - Observation errors are correlated
  - The number of observation is large

Meteosat 9 IR10.8 20080525 0 UTC



RTTOV gen. Meteosat 8 IR10.8 ECMWF Fc 20080525 00 UTC:



Obs and Model (source ECMWF)

Specific issues:

- **Occlusions and coasts** (ocean)
- **2D representation of a 3D observed process** (atmosphere)

Variational data assimilation

Goal : Look for the optimal initial condition, x_0^a, defined as the minimum of the following cost function :

$$J(x) = \underbrace{\frac{1}{2}\|x - x_0^b\|_{\mathbf{B}}^2}_{\text{departure from background}} + \frac{1}{2}\sum_i \underbrace{\|H(\mathcal{M}_{t_0 \to t_i}(x)) - y_i^o\|_{\mathbf{R}}^2}_{\text{misfit to observations}}$$

B : background error covariance matrix
R : observations error covariance matrix

$$\|x\|_{\mathbf{K}}^2 = x^T \mathbf{K}^{-1} x$$

For images: Assuming we can derive a synthetic image q_{t_n} from model output, we can define the distance in different space:

- In the pixel space: $J_o(x_0^a) = \sum_{t_0}^{t_f} \|q_{t_n} - I_{t_n}^{obs}\|_{R_{pix}}^2$
- In the gradient space: $J_o(x_0^a) = \sum_{t_0}^{t_f} \|\nabla q_{t_n} - \nabla I_{t_n}^{obs}\|_{R_\nabla}^2$
- In a wavelet space : $J_o(x_0^a) = \sum_{t_0}^{t_f} \|W q_{t_n} - W I_{t_n}^{obs}\|_{R_W}^2$ where W stands for a wavelet transform.

Outline

Some thoughts about **R**

- Both measurement and representativity errors
- for images both are correlated in space (and time, and between channels, and ...).
  - measurement: same device, preprocessing
  - Note: representativity error are always correlated
- Image-type observations are large, therefore **R** is big.
  - Computing issue: for 3D/4DVar, we need its inverse (actually Inverse $\times$ vector)
  - Storage issue: even if sparse, its inverse is not.
  - thinning does not make much sense
- however, images are dense in space, meaning we can use the same tricks as for **B**

Idea for today's presentation: use the sparseness of multi scale decomposition to try to keep the diagonal approximation for **R** alive (still accounting for some spatial correlation)

Some thoughts about **R**

- Both measurement and representativity errors
- for images both are correlated in space (and time, and between channels, and ...).
  - measurement: same device, preprocessing
  - Note: representativity error are always correlated
- Image-type observations are large, therefore **R** is big.
  - Computing issue: for 3D/4DVar, we need its inverse (actually Inverse $\times$ vector)
  - Storage issue: even if sparse, its inverse is not.
  - thinning does not make much sense
- however, images are dense in space, meaning we can use the same tricks as for **B**

Idea for today's presentation: use the sparseness of multi scale decomposition to try to keep the diagonal approximation for **R** alive (still accounting for some spatial correlation)

Remark:

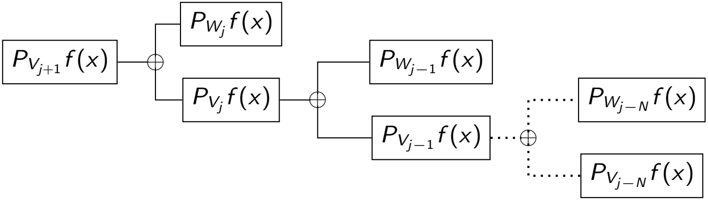
In the following we assume additive gaussian noise while it is often assumed multiplicative

- If lognormal, we are safe (Fletcher and Zupanski, 2006)
- If anything else (e.g. Γ-distribution) go back to Bayes' theorem and scratch your head.

Wavelets

Let V_j the approximation space of a given signal at scale j. One define W_j as the orthogonal complement of V_j :

$$V_{j+1} = V_j \bigoplus W_j$$



Wavelets

Let V_j the approximation space of a given signal at scale j. One define W_j as the orthogonal complement of V_j :

$$V_{j+1} = V_j \bigoplus W_j$$

$$\boxed{P_{V_{j+1}} f(x)} = \boxed{P_{W_j} f(x)} \oplus \boxed{P_{W_{j-1}} f(x)} \oplus \begin{array}{c} \boxed{P_{W_{j-N}} f(x)} \\ \oplus \\ \boxed{P_{V_{j-N}} f(x)} \end{array}$$

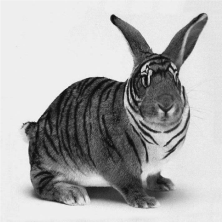
V_{j+1} can be decomposed as:

$$P_{V_{j+1}} f(x) = \underbrace{\sum_k c_{j-N,k} \phi_{j-N,k}(x)}_{\text{Approximation in } v_{j-N}} + \sum_{r=j-N\cdots j} \underbrace{\sum_k d_{r,k} \psi_{r,k}(x)}_{\text{Details in } w_r}$$

V_{j+1} can be decomposed as:

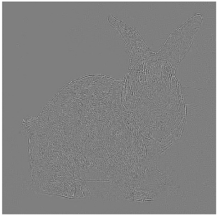
$$P_{V_{j+1}} f(x) = \underbrace{\sum_k c_{j-N,k}\phi_{j-N,k}(x)}_{\text{Approximation in } V_{j-N}} + \sum_{r=j-N\cdots j} \underbrace{\sum_k d_{r,k}\psi_{r,k}(x)}_{\text{Details in } W_r}$$

Original Satellite Image

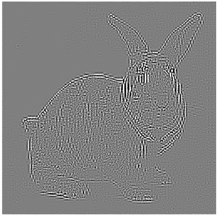


$=$

Details J



Details J-1

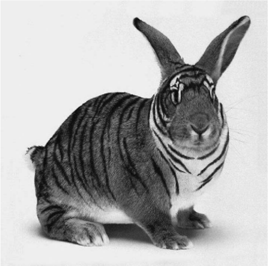


Details J-2

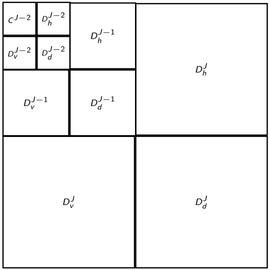
Approximation J-2

Wavelet representation

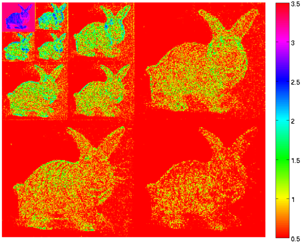
Original image



Wavelet decomposition representation

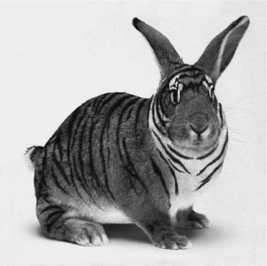


Wavelet coefficients distribution (log)

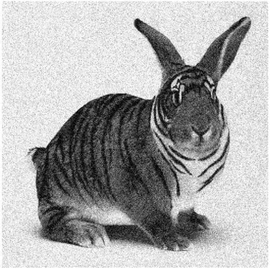


Effect of noise in the wavelet space

Noise free image



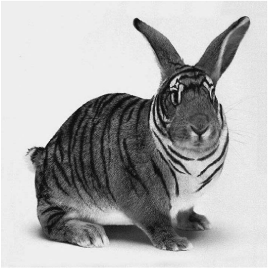
Uncorrelated noise



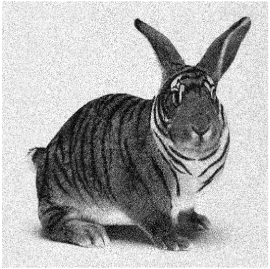
Correlated noise

Effect of noise in the wavelet space

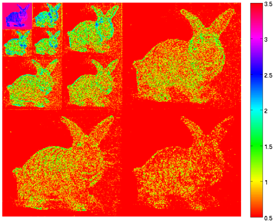
Noise free image



Uncorrelated noise

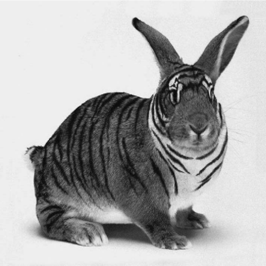


Correlated noise

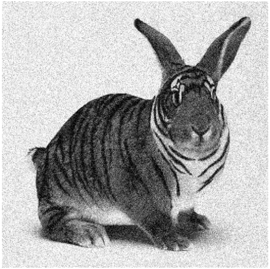


Effect of noise in the wavelet space

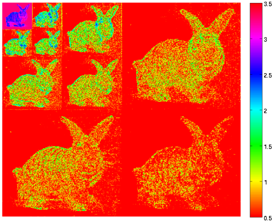
Noise free image

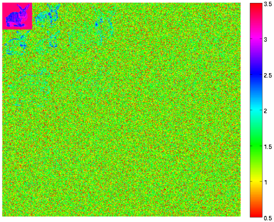


Uncorrelated noise



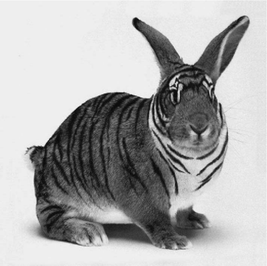
Correlated noise

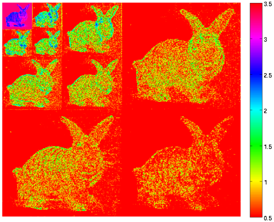




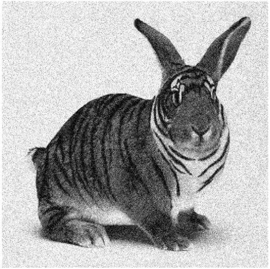
Effect of noise in the wavelet space

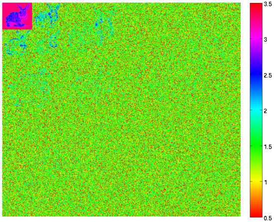
Noise free image



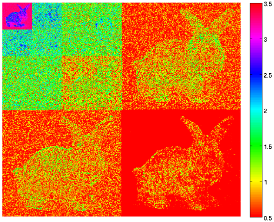


Uncorrelated noise



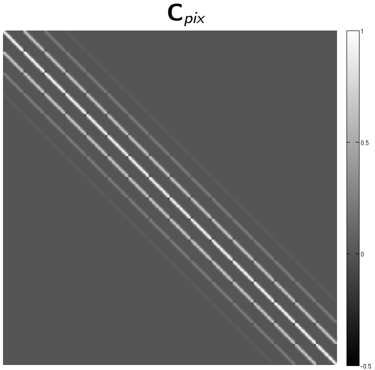


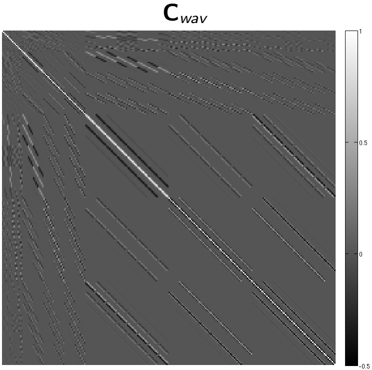
Correlated noise



Correlation matrices

Assuming Gaussian statistics





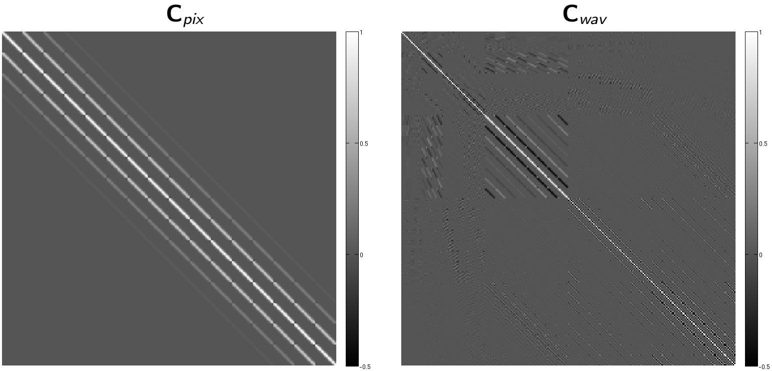
Haar

Significant correlation:

- Between details coefficients at same scale,
- Between details coefficients at different scales,
- Between details and approximation coefficients.

Correlation matrices

Assuming Gaussian statistics

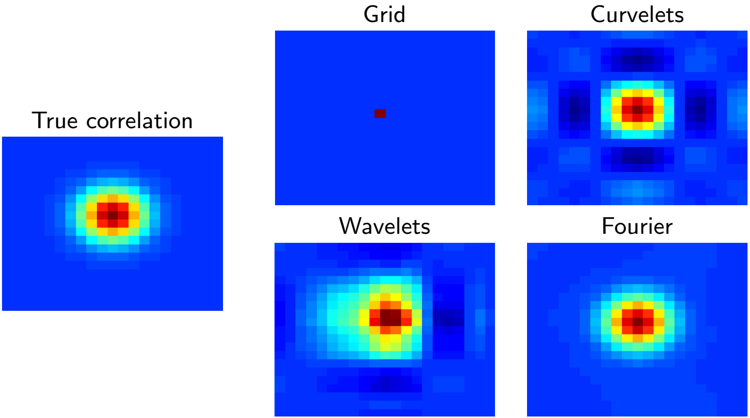


Daubechies

Significant correlation:

- Between details coefficients at same scale,
- Between details coefficients at different scales,
- Between details and approximation coefficients,
- Between coefficients corresponding to image boundaries

Diagonal Approximation for R_{wav}

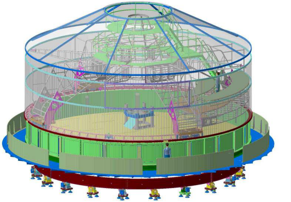


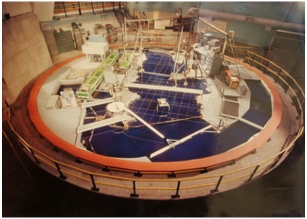
Outline

General framework

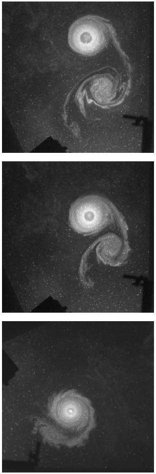
Context and motivations: drift of a vortex [Flór and Eames, 2002]

Coriolis Turntable - LEGI, Grenoble

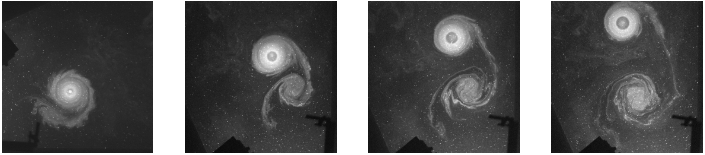




Isolated vortex simulation $\Rightarrow$



Direct Image (Sequence) Assimilation



J.-B. Flór and I. Eames, 2002

shallow-water model for $(\mathbf{u}, \mathbf{v}, \mathbf{h})$

$$(\mathcal{M}) \begin{cases} \partial_t u - u\partial_x u + v\partial_y u - fv + g\partial_x h + \mathcal{D}(u) & = & 0 \\ \partial_t v + u\partial_x v + v\partial_y v + fu + g\partial_y h + \mathcal{D}(v) & = & 0 \\ \partial_t h + \partial_x(hu) + \partial_y(hv) & = & 0 \end{cases}$$

Observation: $\mathbf{y} = \mathcal{T}(\textbf{image})$

$\mathcal{T} = I$: **pixel**
$\mathcal{T} = W$: **multi-scale decomposition**
$\mathcal{T} = \nabla$: **gradient**

Observation operator:
Passive tracer advection
$\mathbf{q}$ represents a synthetic image

$$\begin{cases} \partial_t \mathbf{q} + u\partial_x \mathbf{q} + v\partial_y \mathbf{q} - \nu_T \Delta \mathbf{q} = 0 \\ \mathbf{q}(0) = \mathbf{f}(0) \quad : \quad \text{initial image} \\ (u, v) \quad : \quad \text{verifies } (\mathcal{M}) \end{cases}$$

$$\boxed{\mathcal{H}_{\mathcal{I}}(\mathbf{x}) = \mathcal{T}(\mathbf{q})}$$

Assimilation process

Cost function

$$J(\mathbf{x}) = \frac{\alpha}{2} \parallel \mathbf{x} - \mathbf{x}^b \parallel^2_{\mathbf{B}^{-1}} + \underbrace{\frac{1}{2}\int_0^T \parallel \mathcal{H}_{\mathcal{I}}[\mathbf{x}(t)] - \mathbf{y} \parallel^2_{\mathbf{R}_{\mathcal{I}}^{-1}} dt}_{\text{image - model discrepancy}}$$

Optimisation

- Background at rest ($\mathbf{x^b} = (\mathbf{0}, \mathbf{0}, \mathbf{h_{mean}})^\mathbf{T}$)
- Usual $\mathbf{B}^{1/2}$ change of variable with correlation built using Weaver and Courtier (2001) approach and.
- Progressive (or quasi-static) minimisation technique [Luong et al.(1998), Pires et al. (1996)]
- Minimizer: M1QN3 [Gilbert and Lemarechal]
- twin experiments...

Impact of error statistics

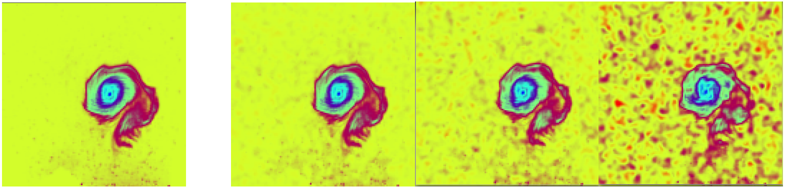


Figure : Signal to Noise Ratio of (from left to right) 26.8 dB, 20.8 dB and 14.8 dB.

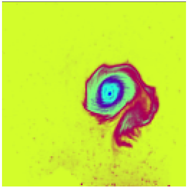
Covariance matrices R are huge and need to be inverted their diagonal restriction are used in practice. In the pixel space and the gradient space this restriction leads to a matrix proportional to the identity:

$$R_{pix}^{scal} = Diag(R_{pix}^{scal}) = \sigma^2 \mathrm{I}_n \quad R_{\nabla}^{scal} = Diag(R_{\nabla}^{scal}) = \tilde{\sigma}^2 \mathrm{I}_{2n}.$$

For wavelets we can use the diagonal restriction of the true covariance matrix in the wavelet space.

$$R_W^{diag} = Diag(R_w) \neq \sigma^2 \mathrm{I}_n$$

Impact of error statistics



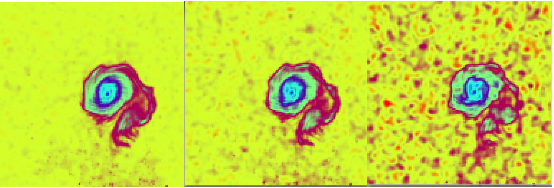
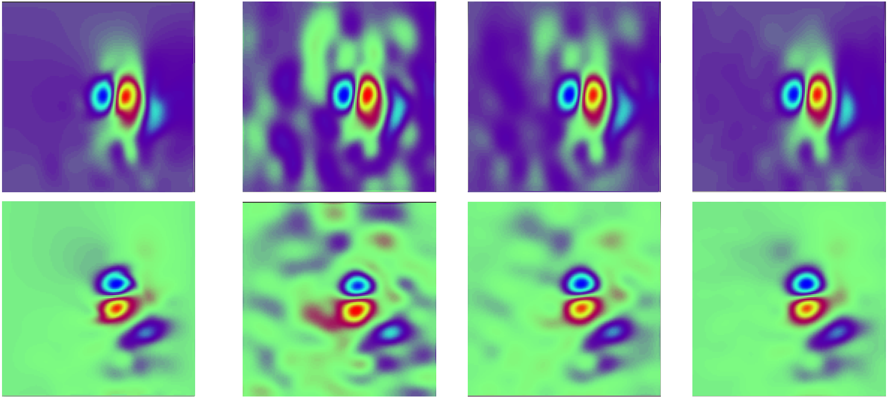


Figure : Signal to Noise Ratio of (from left to right) 26.8 dB, 20.8 dB and 14.8 dB.

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

Table : Mean over 10 experiments of the residual error(in percent).

Results

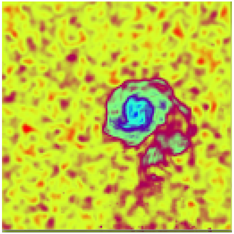


True velocities | Pixel | Gradient | Wavelet Daubechies

Outline

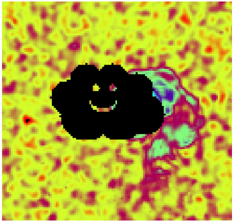
Managing occultations

Many images may suffer from occultation, as for example ocean colour masked by a passing cloud.



Managing occultations

Many images may suffer from occultation, as for example ocean colour masked by a passing cloud.



Managing occultations

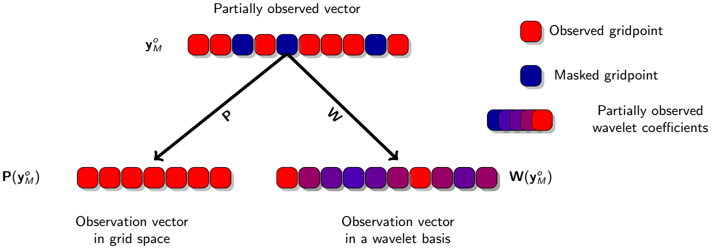


Figure : Working in a wavelet space when the image is partially observed lead to work in larger space (but bearing the same information).

in grid space, it is quite easy to handle

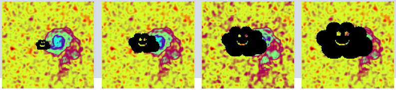
$$\|H(x) - y^{seen}\|^2_{R^{seen}_{grid}}$$

Where H includes a projection (masking).
But in Wavelet space it is more tricky

- naive approach: mask out the synthetic observation and take it as normal image.
- not so naive approach: account for the information content of each wav coefficients

Managing occultations (2)

The more holes, the less Emmental cheese. . .



wavelet decomposition:

$$c^{j-1}[n] = \sum_p h[p-2n]c^j[p]$$

$$d^{j-1}[n] = \sum_p g[p-2n]d^j[p]$$

some attempt to account for missing data:

$$\tilde{\sigma}^2_{d_i} = (\sigma^2_{d_i} + \alpha_{d_i}\sigma^2_{mean}) \times I$$

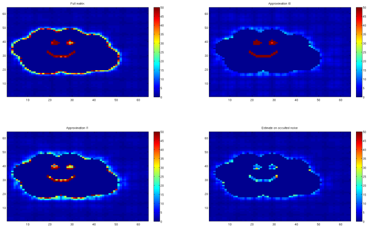
where

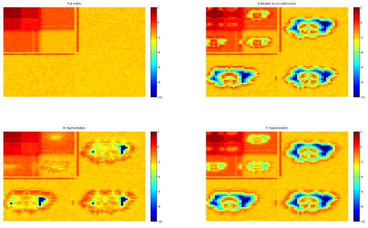
$$\alpha_{d_i} = \left| \frac{\sum g^{occ}[p-2n]}{\sum |g^{occ}[p-2n]|} \right|$$

I^0 number of observed grid point
I^1 weighted number of observed grid point

Conflicting issues:

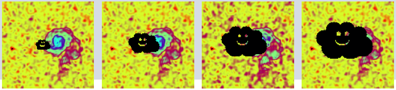
- missing data $\Longrightarrow$ no error
  $\Longrightarrow$ deflate error statistics
- missing data $\Longrightarrow$ more discontinuity in the signal
  $\Longrightarrow$ perturbed small scale coefficients
  $\Longrightarrow$ inflate error statistics





Managing occultations (2)

The more holes, the less Emmental cheese. . .



Conflicting issues:

- missing data $\Longrightarrow$ no error
  $\Longrightarrow$ deflate error statistics
- missing data $\Longrightarrow$ more discontinuity in the signal
  $\Longrightarrow$ perturbed small scale coefficients
  $\Longrightarrow$ inflate error statistics

wavelet decomposition:

$$c^{j-1}[n] = \sum_p h[p-2n]c^j[p]$$

$$d^{j-1}[n] = \sum_p g[p-2n]d^j[p]$$

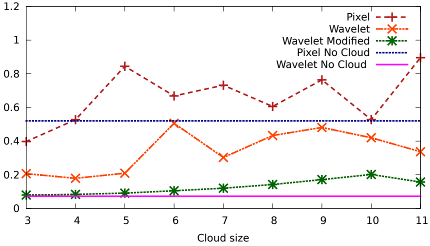
some attempt to account for missing data:

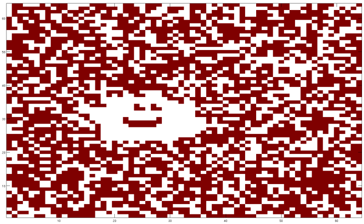
$$\tilde{\sigma}^2_{d_i} = (\sigma^2_{d_i} + \alpha_{d_i}\sigma^2_{mean}) \times I$$

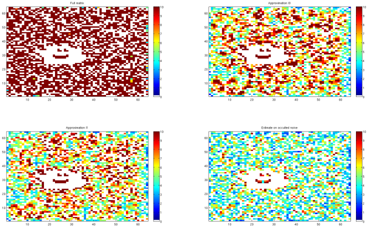
where
$$\alpha_{d_i} = \left| \frac{\sum g^{occ}[p-2n]}{\sum |g^{occ}[p-2n]|} \right|$$

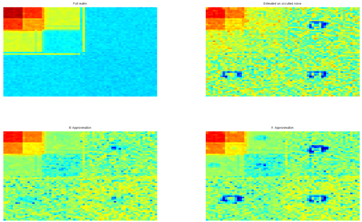
I^0 number of observed grid point
I^1 weighted number of observed grid point

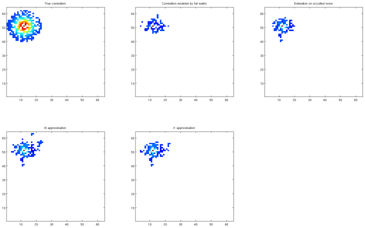
normalized RMS error in meridional velocity

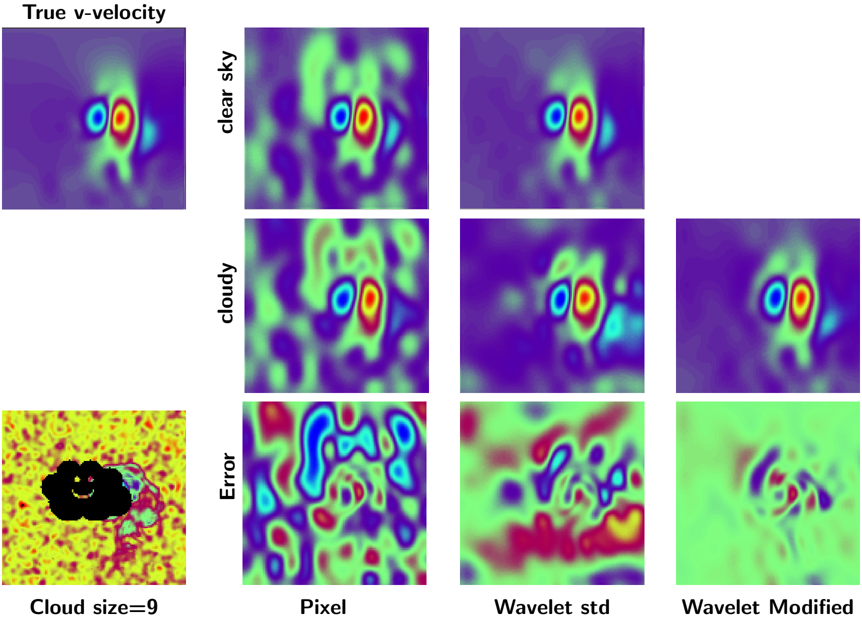











True v-velocity
clear sky
cloudy
Error
Cloud size=9
Pixel
Wavelet std
Wavelet Modified

Outline

Perspective 1: finer representation or more exotic correlation structure

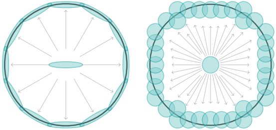
Curvelet decomposition: more adapted to curve discontinuities

- Multi-scale multi orientation transform
- Decomposition: $\mathcal{C}(\mathbf{f}) = \sum_{j,k,l} \langle \mathbf{f}, \varphi_{j,l,k} \rangle \varphi_{j,l,k}$

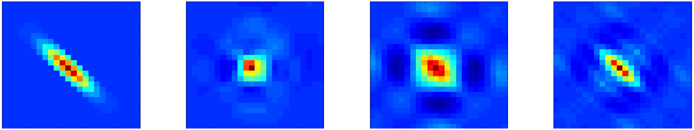
  j: scale; l: orientation; k: position.
- Fast Discrete Curvelet Transf.: $O(n \log n)$
- Adjoint of the decomposition = pseudo-inverse decomposition (reconstruction)

(E. J. Candès and D. L. Donoho, 2004), (L. Demanet 2006), www.curvelet.org

Curvelet atom



Curvelets *Wavelets*



Anisotropic correlation representation from left to right: original, wavelets, curvelets and improved curvelets

Curvelets are more difficult (and expensive) to handle but can produce more fancy correlation

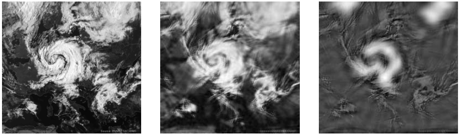
Perspective 2: Multiscale? Oh wait !

The multi-scale aspect can be used for extracting structure from images

$$\mathbf{y} = \tau \circ \mathcal{T}(\textbf{image})$$

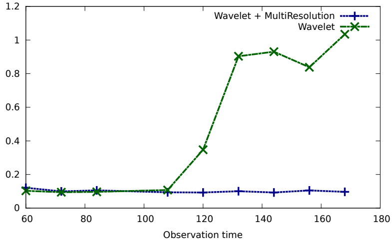
with τ a thresholding operator

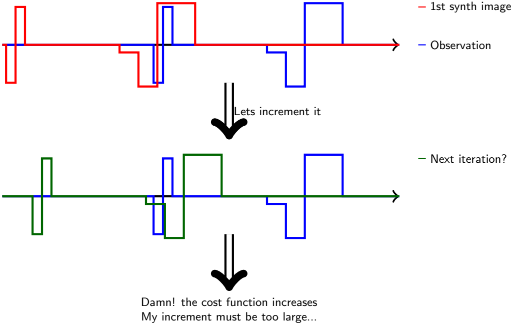
(I. Souopgui's PhD thesis)



(a) Original image (b) Strong threshold (c) Strong scale by scale threshold

It can also be used in a sort of multi-scale framework for the minimisation.





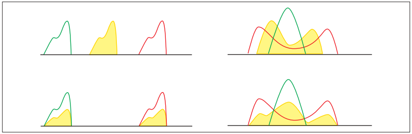
Perspective 3: Optimal Transport
Wasserstein distance

After a feeble attempt to use optical flow output as a metric, we are looking into Wasserstein distances :

$$W_2^2(\rho_0, \rho_1) = \inf_{M \text{ fulfil } (E)} \int |M(x) - x|^2 \rho_0(x)\, dx$$

where (E) is the following non linear constraint

$$(E): \quad \det(\nabla M(x))\rho_1(M(x)) = \rho_0(x)$$



$$J(x) = \frac{1}{2}\|x - y\| + \frac{1}{2}\|x - x^b\|$$

Issues (currently addressed in the Tommi project):

- We need to differentiate this if we want to compute a gradient.
- It is quite expensive to compute.
- Other distances may also be well suited ($\mathcal{H}^{-1}$, ...)
- The mass may not be se same in the observation and its model equivalent
- Have to deal with obstacle (coasts)