Optimization for Bayesian Estimation. The case of Variational Assimilation of Meteorological Observations

Olivier Talagrand & Mohamed Jardak

Laboratoire de Météorologie Dynamique École Normale Supérieure Paris, France

Recent Advances on Optimization

Toulouse, France, July 24 2013



O. Talagrand & M. Jardak





1

200

Optimization for Bayesian Estimation

Purpose of assimilation : reconstruct as accurately as possible the state of the atmospheric or oceanic flow, using all available appropriate information. The latter essentially consists of

- The observations proper, which vary in nature, resolution and accuracy, and are distributed more or less regularly in space and time.
- The physical laws governing the evolution of the flow, available in practice in the form of a discretized, and necessarily approximate, numerical model.
- 'Asymptotic' properties of the flow, such as, e. g., geostrophic balance of middle latitudes.
 Although they basically are necessary consequences of the physical laws which govern the flow, these properties can usefully be explicitly introduced in the assimilation process.

Both observations and 'model' are affected with some uncertainty \Rightarrow uncertainty on the estimate.

For some reason, uncertainty is conveniently described by probability distributions (don't know too well why, but it works).

Assimilation is considered here as a problem in bayesian estimation.

Determine the conditional probability distribution for the state of the system, knowing everything we know.

Data of the form

 $z = \Gamma x + \zeta, \qquad \qquad \zeta \sim \mathcal{N}[\mu, S]$

Known data vector z belongs to *data space* \mathcal{D} , $dim\mathcal{D} = m$, Unknown state vector x belongs to *state space* \mathcal{X} , $dim\mathcal{X} = n$ Γ known (*mxn*)-matrix, ζ unknown 'error'

Then conditional probability distribution is

 $P(x \mid z) = \mathcal{N}[x^a, P^a]$

where

$$x^{a} = (\Gamma^{T} S^{-1} \Gamma)^{-1} \Gamma^{T} S^{-1} [z - \mu]$$
$$P^{a} = (\Gamma^{T} S^{-1} \Gamma)^{-1}$$

Determinacy condition : rank $\Gamma = n$. Requires $m \ge n$.

Variational form.

Conditional expectation x^a minimizes following scalar *objective function*, defined on state space X

$\boldsymbol{\xi} \in \boldsymbol{\mathcal{X}} \to \boldsymbol{\mathcal{J}}(\boldsymbol{\xi}) = (1/2) \left[\boldsymbol{\varGamma}\boldsymbol{\xi} - (\boldsymbol{z} - \boldsymbol{\mu})\right]^{\mathrm{T}} S^{-1} \left[\boldsymbol{\varGamma}\boldsymbol{\xi} - (\boldsymbol{z} - \boldsymbol{\mu})\right]$

Variational assimilation, as described by previous lecturers, and implemented heuristically in many places on (not too) nonlinear data operators Γ .

Conditional probability distribution is

$$P(x \mid z) = \mathcal{N}[x^a, P^a]$$

with

$$x^{a} = (\Gamma^{T} S^{-1} \Gamma)^{-1} \Gamma^{T} S^{-1} [z - \mu]$$
$$P^{a} = (\Gamma^{T} S^{-1} \Gamma)^{-1}$$

Ready recipe for determining Monte-Carlo sample of conditional pdf $P(x \mid z)$:

- Perturb data vector z according to its own error probability distribution

 $z \rightarrow z^{*} = z + \delta, \qquad \delta \sim \mathcal{N}[0, S]$

and compute

$$x^{\prime a} = (\Gamma^{\mathrm{T}} S^{-1} \Gamma)^{-1} \Gamma^{\mathrm{T}} S^{-1} [z^{\prime} - \mu]$$

 x^{a} is distributed according to $\mathcal{N}[x^{a}, P^{a}]$

Ensemble Variational Assimilation (EnsVar) implements that algorithm, the expectations x^{a} being computed by standard variational assimilation (optimization)

EnsVar results : the linearized Lorenz96 model (5 days)





▲□▶ ▲圖▶ ▲圖▶ ▲圖▶

1

SQC

O. Talagrand & M. Jardak

Optimization for Bayesian Estimation

EnsVar results : consistency

Consistency : under linearity the expectation of the objective function \mathfrak{J} at its minimum is half the number of observations p, $\mathbb{E}(\mathfrak{J}(x_{opt})) = \frac{p}{2}$ and under Gaussianity we have $\mathbb{V}ar(\mathfrak{J}(x_{opt})) = p$.



▲□▶▲□▶▲目▶▲目▶ 目 のQで

Purpose of the present work

- Objectively evaluate EnsVar as a probabilistic estimator in nonlinear and/or non-Gaussian cases.

- Objectively compare with other existing ensemble assimilation algorithms : *Ensemble Kalman Filter (EnKF), Particle Filters (PF)*

- Simulations performed on two small-dimensional chaotic systems, the Lorenz'96 model and the Kuramoto-Sivashinsky equation

How to objectively evaluate the performance of an ensemble (or more generally probabilistic) estimation system ?

- There is no general objective criterion for Bayesianity

- We use instead the weaker property of *reliability*, *i. e.* statistical consistency between predicted probabilities and observed frequencies of occurrence (it rains with frequency 40% in the circonstances where I have predicted 40% probability for rain).

Reliability can be objectively validated, provided a large enough sample of realizations of the estimation system is available.

Bayesianity implies reliability, the converse not being true.

The Lorenz96 model

• Forward model

$$\frac{dx_k}{dt} = (x_{k+1} - x_{k-2})x_{k-1} - x_k + F \quad \text{for } k = 1, \cdots, N$$

- Set-up parameters :
 - **1** the index k is cyclic so that $x_{k-N} = x_{k+N} = x_k$.
 - 2 F = 8, external driving force.
 - $\bigcirc -x_k$, a damping term.
 - \bigcirc N = 40, the system size.
 - \bigcirc Nens = 30, number of ensemble members.
 - $\frac{1}{\lambda_{max}} \simeq 2.5 days, \lambda_{max}$ the largest Lyapunov exponent.
 - $\Delta t = 0.05 = 6hours$, the time step.
 - If frequency of observations : every 12 hours.
 - number of realizations : 9000 realizations.

▲□ > ▲ 目 > ▲ 目 > ● ■ ● ● ● ●

EnsVar : the non-linear Lorenz96 model (4.5 days $\simeq 1$ TU)







DQC

O. Talagrand & M. Jardak

Optimization for Bayesian Estimation

EnsVar : the non-linear Lorenz96 model (10 days \simeq 2 TU)



EnsVar : consistency



- イロト (個) (目) (目) ヨー のへで

EnsVar results : Objective function section



O. Talagrand & M. Jardak Optimization for Bayesian Estimation

Quasi-Static Variational Assimilation (QSVA)



EnsVar : the non-linear Lorenz96 model 10 days with **QSVA**



O. Talagrand & M. Jardak

Optimization for Bayesian Estimation

-2

0.8

-2

▶ < ≣ ▶

1

9QC

0

EnsVar : the non-linear Lorenz96 model 18 days with **QSVA**



threshold



=

.

< ≣

Ŧ

DQC

O. Talagrand & M. Jardak

Optimization for Bayesian Estimation

EnsVar : observation frequency impact

Impact on the reliability and resolution



- Results are independent of the Gaussian character of the observation errors (trials have been made with various parobability distributions, in particular Cauchy distributions, which do not have a finite variance)
- Ensembles produced by EnsVar are very close to Gaussian, even in strongly nonlinear cases.

EnKF results : the non-linear Lorenz96 model

Nonlinear case : 10 days time length

elabily dama

D4 0.5 D3 pendicied probability









▲□▶★圖▶★≣▶★≣▶ ≣ の�?

O. Talagrand & M. Jardak Optimization for Bayesian Estimation

PF results : the non-linear Lorenz96 model

Nonlinear case : 10 days time length











- ◆ロ> ◆聞> ◆ヨ> ◆ヨ> 三三 のへで

EnsVar : 5 days forecast



▲口▶▲舂▶▲≧▶▲≧▶ ≧ の�?

EnKF : 5days forecast



Reference & ensemble member solutions at the end of the forecast Reference & ensemble member solutions

O. Talagrand & M. Jardak Optimization for Bayesian Estimation

SAC

1

PF : 5days forecast



Weak constraint EnsVar

• define the objective function.

$$\mathfrak{J}(x,\eta_1,\eta_2,\cdots,\eta_{N-1},\eta_N) = \frac{1}{2} \left\{ (x-x_b)^T \mathbf{B}^{-1} (x-x_b) \right\} +$$

$$\frac{1}{2}\sum_{i=0}^{N}\left\{(y_i - H_i(x_i))^T \mathbf{R}_i^{-1}(y_i - H_i(x_i))\right\} + \frac{1}{2}\sum_{i=1}^{N}\eta_i^T \mathbf{Q}_i^{-1}\eta_i$$

- B background error covariance matrix and R observation error covariance matrix.
- Q model error covariance matrix.
- $I : \mathbb{R}^{state} \longrightarrow \mathbb{R}^{obs} \text{ observation operator.}$
- x_b background state vector and y_i observation vector at time $t = t_i$.
- $\textbf{ 0 } \eta_i \text{ model error vector at } t = t_i \text{ with } x(t_i) = \mathfrak{M}_{t_i \leftarrow t_{i-1}}(x(t_{i-1})) + \eta_i$
- find the optimal control variable $(x_0^{opt}, \eta_1^{opt}, \eta_2^{opt}, \cdots, \eta_N^{opt})$ and the optimal trajectory x^{opt} .

$$(x_0^{opt}, \eta_1^{opt}, \eta_2^{opt}, \cdots, \eta_N^{opt}) = \min_{x, \eta_1, \eta_2, \cdots, \eta_N \in \mathfrak{A}} \mathfrak{J}(x, \eta_1, \eta_2, \cdots, \eta_N)$$

$$x_i^{opt} = \mathfrak{M}_{t_i \leftarrow t_{i-1}} (\mathfrak{M}_{t_{i-1} \leftarrow t_{i-2}} \cdots (\mathfrak{M}_{t_2 \leftarrow t_1} (\mathfrak{M}_{t_1 \leftarrow t_0} (x_0^{opt}) + \eta_1^{opt}) + \eta_2^{opt}) \cdots + \eta_{i-1}^{opt}) + \eta_i^{opt} = \mathfrak{M}_{t_i \leftarrow t_{i-2}} \cdots (\mathfrak{M}_{t_2 \leftarrow t_1} (\mathfrak{M}_{t_1 \leftarrow t_0} (x_0^{opt}) + \eta_1^{opt}) + \eta_2^{opt}) \cdots + \eta_i^{opt}) + \mathfrak{M}_{t_i \leftarrow t_{i-2}} \cdots (\mathfrak{M}_{t_2 \leftarrow t_1} (\mathfrak{M}_{t_1 \leftarrow t_0} (x_0^{opt}) + \eta_1^{opt}) + \eta_2^{opt}) \cdots + \eta_i^{opt}) + \mathfrak{M}_{t_i \leftarrow t_{i-2}} \cdots (\mathfrak{M}_{t_2 \leftarrow t_1} (\mathfrak{M}_{t_1 \leftarrow t_0} (x_0^{opt}) + \eta_1^{opt}) + \eta_2^{opt}) \cdots + \eta_i^{opt}) + \mathfrak{M}_{t_i \leftarrow t_{i-2}} \cdots (\mathfrak{M}_{t_2 \leftarrow t_1} (\mathfrak{M}_{t_1 \leftarrow t_0} (x_0^{opt}) + \eta_1^{opt}) + \mathfrak{M}_{t_i \leftarrow t_{i-2}} \cdots (\mathfrak{M}_{t_i \leftarrow t_0} (x_0^{opt}) + \eta_1^{opt}) + \mathfrak{M}_{t_i \leftarrow t_0} (\mathfrak{M}_{t_i \leftarrow t_0} (x_0^{opt}) + \mathfrak{M}_{t_i \leftarrow t_0} (x_0^{opt}) + \mathfrak{M}_{t_i \leftarrow t_0} (\mathfrak{M}_{t_i \leftarrow t_0} (x_0^{opt}) + \mathfrak{M}_{t_i \leftarrow t_0} (\mathfrak{M}_{t_i \leftarrow t_0} (x_0^{opt}) + \mathfrak{M}_{t_i \leftarrow t_0} (x_0^{opt}) + \mathfrak{M}_{t_i \leftarrow t_0} (\mathfrak{M}_{t_i \leftarrow t_0} (x_0^{opt}) + \mathfrak{M}_{t_i \leftarrow t_0} (x_0^{opt}) + \mathfrak{M}_{t_i \leftarrow t_0} (\mathfrak{M}_{t_i \leftarrow t_0} (x_0^{opt}) + \mathfrak{M}_{t_i \leftarrow t_0} (\mathfrak{M}_{t_i \leftarrow t_0} (x_0^{opt}) + \mathfrak{M}_{t_i \leftarrow t$$

Weak EnsVar : the Lorenz96 model 8 days



O. Talagrand & M. Jardak Optimization for Bayesian Estimation

200

Summary

- Under non-linearity and non-Gaussianity the EnsVar is a reliable and consistent ensemble estimator (provided the QSVA is used for long DA windows).
- EnsVar is at least as good an estimator as EnKF and PF.
- Similar results have been obtained for the Kuramuto-Sivashinsky model.

Pros

- Easy to implement when having a 4D-Var code
- Highly parallelizable
- No problems with algorithm stability (i.e. no ensemble collapse, no need for localization and inflation, no need for weight resampling)
- Propagates information in both ways and takes into account temporally correlated errors

Cons

- Costly (Nens 4D-Var assimilations).
- Empirical.
- Cycling of the process (work in progress).

Thank You

O. Talagrand & M. Jardak Optimization for Bayesian Estimation