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
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

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

Known data vector \( z \) belongs to \textit{data space} \( D \), \( \dim D = m \),
Unknown state vector \( x \) belongs to \textit{state space} \( X \), \( \dim X = n \)
\( \Gamma \) known \((mxn)\)-matrix, \( \zeta \) unknown ‘error’

Then conditional probability distribution is

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

where

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

\textit{Determinacy condition} : \( \text{rank} \Gamma = n \). Requires \( m \geq n \).
Variational form.

Conditional expectation $x^a$ minimizes following scalar \textit{objective function}, defined on state space $X$

$$\xi \in X \rightarrow J(\xi) = (1/2) \left[ \Gamma \xi - (z-\mu) \right]^T S^{-1} \left[ \Gamma \xi - (z-\mu) \right]$$

\textit{Variational assimilation}, as described by previous lecturers, and implemented heuristically in many places on (not too) nonlinear data operators $\Gamma$. 
Conditional probability distribution is

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

with

\[
\begin{align*}
x^a &= (\Gamma^T S^{-1} \Gamma)^{-1} \Gamma^T S^{-1} [z - \mu] \\
P^a &= (\Gamma^T S^{-1} \Gamma)^{-1}
\end{align*}
\]

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

\[ z \to z' = z + \delta, \quad \delta \sim \mathcal{N}[0, S] \]

and compute

\[
x'^a = (\Gamma^T S^{-1} \Gamma)^{-1} \Gamma^T S^{-1} [z' - \mu]
\]

\( x'^a \) 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)

- Ensemble optimal control, reference and observations
- Ensemble optimal trajectories and reference solutions
- Rank histogram
- Reliability diagram
- Brier skill scores
- Errors
Consistency: under linearity the expectation of the objective function $\mathcal{J}$ at its minimum is half the number of observations $p$, $\mathbb{E}(\mathcal{J}(x_{opt})) = \frac{p}{2}$ and under Gaussianity we have $\text{Var}(\mathcal{J}(x_{opt})) = p$. 
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 circumstances 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

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

- 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.
  3. $-x_k$, a damping term.
  4. $N = 40$, the system size.
  5. $N_{ens} = 30$, number of ensemble members.
  6. $\frac{1}{\lambda_{max}} \simeq 2.5 \text{days}$, $\lambda_{max}$ the largest Lyapunov exponent.
  7. $\Delta t = 0.05 = 6 \text{hours}$, the time step.
  8. frequency of observations: every 12 hours.
  9. number of realizations: 9000 realizations.
EnsVar: the non-linear Lorenz96 model (4.5 days $\sim 1$ TU)
EnsVar: the non-linear Lorenz96 model (10 days \(\sim 2\) TU)
Ensemble Variance: consistency
EnsVar results: Objective function section

The graph shows the objective function section and an interpolated quadratic function. The objective function section is represented by a black line, while the interpolated quadratic is shown in red.

The x-axis represents the variable $\lambda$, and the y-axis represents $\Delta \lambda_{\text{tot}} + (1-\lambda) \lambda_{\text{min}}$, scaled by $10^5$. The range of $\lambda$ is from -3 to 4, and the range of the y-axis is from 0 to $8 \times 10^5$.
Quasi-Static Variational Assimilation (QSVA)

Data Assimilation over $[0 \ T]$ with $T = N \ dt = M \ d\tau$

4D-Var over $[0 \ \tau]$ starting from the observations

4D-Var over $[0 \ 2\tau]$ starting from the minimizer found above

Repeat the rule

4D-Var over $[0 \ T]$ starting from the minimizer found above and set the minimum as absolute
EnsVar: the non-linear Lorenz96 model 10 days with QSVA
EnsVar: the non-linear Lorenz96 model 18 days with QSVA

O. Talagrand & M. Jardak

Optimization for Bayesian Estimation
EnsVar: observation frequency impact

Impact on the reliability and resolution

- Blue: reliability
- Red: every other time step
- Green: every 4 time steps
- Cyan: every 10 time steps

Threshold vs. reliability & resolution

O. Talagrand & M. Jardak
Optimization for Bayesian Estimation
- Results are independent of the Gaussian character of the observation errors (trials have been made with various probability 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

![Graphs showing ensemble, observations, and ensemble solutions at the end of the simulation.](image)

![Graphs showing ensemble trajectories and ensemble solutions.](image)

![Histograms and velocity diagrams.](image)
PF results: the non-linear Lorenz96 model

Nonlinear case: 10 days time length
EnsVar: 5 days forecast

- Ensemble optimal control, reference and observations
- Ensemble optimal trajectories and reference solutions
- Rank histogram
- Reliability diagram
EnKF: 5 days forecast

Reference & ensemble member solutions at the end of the forecast

Reference & ensemble member solutions

Rank histogram at the end of the forecast

reliability diagram

O. Talagrand & M. Jardak

Optimization for Bayesian Estimation
PF: 5 days forecast

Reference & particle filter solutions at the end of the forecast

Reference & particle filter solutions

Rank histogram at the end of the forecast

Reliability diagram

O. Talagrand & M. Jardak
Optimization for Bayesian Estimation
Weak constraint EnsVar

- define the objective function.

\[
\mathcal{J}(x, \eta_1, \eta_2, \cdots, \eta_{N-1}, \eta_N) = \frac{1}{2} \left\{ (x - x_b)^T B^{-1} (x - x_b) \right\} + \\
\frac{1}{2} \sum_{i=0}^{N} \left\{ (y_i - H_i(x_i))^T R_i^{-1} (y_i - H_i(x_i)) \right\} + \frac{1}{2} \sum_{i=1}^{N} \eta_i^T Q_i^{-1} \eta_i
\]

1. B background error covariance matrix and R observation error covariance matrix.

2. Q model error covariance matrix.

3. \( H : \mathbb{R}^{\text{state}} \rightarrow \mathbb{R}^{\text{obs}} \) observation operator.

4. \( x_b \) background state vector and \( y_i \) observation vector at time \( t = t_i \).

5. \( \eta_i \) model error vector at \( t = t_i \) with \( x(t_i) = M_{t_i \leftarrow t_{i-1}}(x(t_{i-1})) + \eta_i \)

- find the optimal control variable \((x_0^{\text{opt}}, \eta_1^{\text{opt}}, \eta_2^{\text{opt}}, \cdots, \eta_N^{\text{opt}})\) and the optimal trajectory \( x^{\text{opt}} \).

\[
(x_0^{\text{opt}}, \eta_1^{\text{opt}}, \eta_2^{\text{opt}}, \cdots, \eta_N^{\text{opt}}) = \min_{x, \eta_1, \eta_2, \cdots, \eta_N \in \Omega} \mathcal{J}(x, \eta_1, \eta_2, \cdots, \eta_N)
\]

\[
x_i^{\text{opt}} = M_{t_i \leftarrow t_{i-1}} (M_{t_{i-1} \leftarrow t_{i-2}} \cdots M_{t_2 \leftarrow t_1} (M_{t_1 \leftarrow t_0} (x_0^{\text{opt}} + \eta_1^{\text{opt}}) + \eta_2^{\text{opt}}) + \cdots + \eta_{i-1}^{\text{opt}}) + \eta_i^{\text{opt}}
\]
Weak EnsVar: the Lorenz96 model 8 days
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 Kuramoto-Sivashinsky model.
EnsVar: Pros and cons

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