Joint Estimation of Model and Observation Error Covariance Matrices in Data Assimilation: a Review

Pierre Tandeo

Associate professor at IMT Atlantique, Brest, France

27th of September 2018, Rennes, France
Before starting

Why a review paper?

AMS Journals Welcome Review Articles

David M. Schultz
Chair, Subcommittee on Reviews, and Chief Editor, Monthly Weather Review

See all authors & affiliations

https://doi.org/10.1175/MWR-D-18-0114.1
Published Online: 23 April 2018

Works in collaboration with:

P. Ailliot
Univ. Brest
France

M. Bocquet
CEREA
France

A. Carrassi
NERSC
Norway

T. Miyoshi
RIKEN
Japan

M. Pulido
Univ. Reading
UK

Y. Zhen
IMT A
France
State-space model

Nonlinear formulation with additive and Gaussian noises:

\[
\begin{align*}
\mathbf{x}(t) &= \mathcal{M}(\mathbf{x}(t - dt)) + \eta(t) \\
\mathbf{y}(t) &= \mathcal{H}(\mathbf{x}(t)) + \epsilon(t)
\end{align*}
\]  

(1)  
(2)

with:

- \(\mathcal{M}\) the dynamical model (physical or statistical)
- \(\mathcal{H}\) the transformation from state \(\mathbf{x}\) to observations \(\mathbf{y}\)
- \(\eta(t) \sim \mathcal{N}(0, Q(t))\) the model error
- \(\epsilon(t) \sim \mathcal{N}(0, R(t))\) the observation error

⇒ Estimate both \(Q\) and \(R\) is a key point in DA
Importance of errors

Simple univariate, linear and Gaussian state-space model:

\[
\begin{align*}
    x(k) &= 0.95x(k - 1) + \eta(k) \\
    y(k) &= x(k) + \epsilon(k)
\end{align*}
\]

with \( \eta(k) \sim \mathcal{N}(0, Q^t = 1) \) and \( \epsilon(k) \sim \mathcal{N}(0, R^t = 1) \)
Importance of errors (bad $Q/R$ ratio, underest. of $Q$ or $R$)

- **Bad $Q/R$ ratio:**
  - (a) $Q/R = 0.1$, $Q$ too low, too confident on the model
  - (b) $Q/R = 10$, $R$ too low, too confident on the observations

- Impact on **state reconstruction** (RMSE)
Importance of errors (bad $Q/R$ ratio, overest. of $Q$ or $R$)

- **Bad $Q/R$ ratio:**
  - (c) $Q/R = 10$, same RMSE as (b)
  - (d) $Q/R = 0.1$, same RMSE as (a)
Importance of errors (good $Q/R$ ratio)

Kalman smoother with $Q = 0.1Q^t$ & $R = 0.1R^t$

(e) $Q$ and $R$ too low
(f) $Q$ and $R$ too high

No impact on RMSE

Impact on uncertainty quantification (% in CI)
Link between $||Q||/||R||$ ratio and Kalman equations

with the courtesy of Alberto Carrassi, from [Carrassi et al., 2018]

Kalman filter gain:

$$K(k) = P^f(k)H^\top \left(HP^f(k)H^\top + R(k)\right)^{-1}$$  \hspace{1cm} (3)

with $P^f(k) = M(k)P^a(k-1)M(k)^\top + Q(k)$  \hspace{1cm} (4)
Compensation of bads $Q$ or $R$

Estim. of $Q$ or $R$ (EM algorithm, [Shumway and Stoffer, 1982]):

$\Rightarrow$ Can not compensate for bads $Q$ or $R$!
State of the art

- 4 main families of methods to jointly estimate $Q$ and $R$
- More than 50 papers in DA from the 90’s
- No comparison between all these methods
- Poor links with signal processing and statistical communities

⇒ Estimation of $Q$ and $R$ in DA initiated by [Dee, 1995]
Innovation-based methods

Innovation statistics (mean and covariance):

\[ d(k) = y(k) - Hx^f(k) \]  \tag{5} \\
\[ \Sigma(k) = HP^f(k)H^\top + R(k) \]  \tag{6} 

Ideas:

\begin{itemize}
  \item use various innovations (not only the current one)
  \item ”method of moments” (theoretical=empirical moments)
\end{itemize}

Methods:

\begin{itemize}
  \item I) Innovations in the observation space
  \item II) Lag-innovation statistics
\end{itemize}
I) Innovations in the observation space

Main references:

- [Desroziers et al., 2005] ⇒ use of various innovations among
  \[d^{o-f}(k) = y(k) - Hx^f(k)\] and \[d^{o-a}(k) = y(k) - Hx^a(k)\]
- [Li et al., 2009] ⇒ estimation of covariance \( R \) and covariance inflation for \( P^f \)

Solve the following system:

\[
\begin{align*}
\mathbb{E} \left[ d^{o-f}(k) d^{o-f}(k)^\top \right] &= H P^f(k) H^\top + R(k) = \Sigma(k) \quad (7) \\
\mathbb{E} \left[ d^{o-a}(k) d^{o-f}(k)^\top \right] &= R(k) \quad (8)
\end{align*}
\]

Specific points:

- online approach (adaptive \( Q(k) \) and \( R(k) \))
- no direct estimation of \( Q(k) \) but inflation of \( P^f(k) \)
- low computation cost
II) Lag-innovation statistics

Main references:

- [Berry and Sauer, 2013] ⇒ only lag-1 innovation
- [Zhen and Harlim, 2015] ⇒ comparison lag-1 VS lag-L

Solve the following system (example of lag-0 VS lag-1):

\[
\begin{array}{l}
\mathbb{E} \left[ d(k) d(k) \top \right] = H P^f(k) H \top + R(k) = \Sigma(k) \\
\mathbb{E} \left[ d(k) d(k - 1) \top \right] = H F(k) P^f(k - 1) H \top \\
- H F(k) K(k - 1) \Sigma(k - 1)
\end{array}
\]

Specific points:

- online approach (adaptive $Q(k)$ and $R(k)$)
- sensitive to smoothing parameters
- moderate computation cost
Likelihood-based methods

Innovation likelihood:

\[ p(y(k)|y(1: k-1)) \sim \mathcal{N}(d(k) = y(k) - Hx^f(k), \Sigma(k) = HP^f(k)H^\top + R(k)) \] (11)

Ideas:

- use various innovation likelihoods (not only the current one)
- "likelihood methods" (max. likelihood, max. a posteriori)

Methods:

- III) Bayesian approaches
- IV) Maximization of the total likelihood
III) Bayesian approaches

Main references:
- classic in the statistical community
- [Stroud and Bengtsson, 2007] ⇒ joint parameter for $Q$ and $R$
- [Ueno and Nakamura, 2016] ⇒ parameterization of $R$
- [Stroud et al., 2018] ⇒ spatial parameterization of $R$

Joint distribution of $x$ and $\theta$ (shape param. of $Q$ and $R$):

\[
p (x(k), \theta(k)|y(1:k)) = \\
p (x(k)|\theta(k), y(1:k)) p (\theta(k)|y(1:k))
\]

with $\theta$ following a priori distributions (with given hyperparam.).

Specific points:
- online approach (but non adaptive in practice)
- estimation of both $Q$ and $R$ problematic?
- high computation cost
IV) Maximization of the total likelihood

Main references:

- use the Expectation-Maximization (EM) algorithm
- [Shumway and Stoffer, 1982] ⇒ linear and Gaussian case
- [Tandeo et al., 2015] ⇒ for nonlinear observation operator $\mathcal{H}$
- [Dreano et al., 2017] ⇒ for nonlinear dynamical model $\mathcal{M}$

Maximize the likelihood function:

$$p(\mathbf{y}(1 : K), \mathbf{x}(0 : K)|\mathbf{Q}, \mathbf{R}) = \frac{p(\mathbf{x}(0))}{\prod_{k=1}^{K} p(\mathbf{x}(k)|\mathbf{x}(k - 1), \mathbf{Q})} \prod_{k=1}^{K} p(\mathbf{y}(k)|\mathbf{x}(k), \mathbf{R})$$  \hspace{1cm} (13)

Specific points:

- offline approach (constant $\mathbf{Q}$ and $\mathbf{R}$)
- no additional parameter, robust in practice
- high computation cost (EM iterates filtering/smoothing)
Conclusions

- Joint estimation of $Q$ and $R$ in DA:
  - 4 main methods (innovation-based, likelihood-based)
  - secondary methods (state augmentation, analysis increment, covariance matching, $\chi^2$ test, cross-validation, etc...)

- Review paper:
  - exhaustive bibliography?
  - your feedbacks are welcome!
  - accepted for submission to the MWR

- Available at (https://arxiv.org/abs/1807.11221):

**Joint Estimation of Model and Observation Error Covariance Matrices in Data Assimilation: a Review**

Pierre Tandeo, Pierre Alliot, Marc Bocquet, Alberto Carrassi, Takemasa Miyoshi, Manuel Pulido, Yicun Zhen

(Submitted on 30 Jul 2018)

This paper is a review of a crucial topic in data assimilation: the joint estimation of model $Q$ and observation $R$ matrices. These covariances define the observational and model errors via additive Gaussian white noises in state-space models, the most common way of formulating data assimilation problems. They are crucial because they control the relative weights of the model forecasts and observations in reconstructing the state, and several methods have been proposed since the 90's for their estimation. Some of them are based on the moments of various innovations, including those in the observation space or lag-innovations. Alternatively, other methods use likelihood functions and maximum likelihood estimators or Bayesian approaches. This review aims at providing a comprehensive summary of the proposed methodologies and factually describing them as they appear in the literature. We also discuss (i) remaining challenges for the different estimation methods, (ii) some suggestions for possible improvements and combinations of the approaches and (iii) perspectives for future works, in particular numerical comparisons using toy-experiments and practical implementations in data assimilation systems.
Perspectives

- Combine different approaches:
  - first offline for calibration of $Q$ and $R$...
  - ... and then online for adaptive estimation of $Q(k)$ and $R(k)$
- Possible extensions:
  - estimation of the initial condition (background $x^b$ and $B$)
  - estimation of systematic biases (deterministic part of $\eta$ and $\epsilon$)
- Numerical comparison of methods:
  - using toy models and mid-complexity GCMs
  - using additive, multiplicative and parametric errors
Thank you! Any questions?
Bélanger, P. R. (1974).
Estimation of noise covariance matrices for a linear time-varying stochastic process.

Adaptive ensemble Kalman filtering of non-linear systems.

Data Assimilation in the Geosciences: An overview on methods, issues and perspectives.
*WIREs Clim Change*, page e535.

On-line Estimation of Error Covariance Parameters for Atmospheric Data Assimilation.

Diagnosis of observation, background and analysis-error statistics in observation space.

Estimating model error covariances in nonlinear state-space models using Kalman smoothing and the expectation-maximisation algorithm.

Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter.

On the Identification of Variances and Adaptive Kalman Filtering.

An Approach to Time Series Smoothing and Forecasting using the EM Algorithm.
Sequential State and Variance Estimation within the Ensemble Kalman Filter.

A Bayesian adaptive ensemble Kalman filter for sequential state and parameter estimation.

Offline parameter estimation using EnKF and maximum likelihood error covariance estimates: Application to a subgrid-scale orography parametrization.

Bayesian estimation of the observation-error covariance matrix in ensemble-based filters.

Adaptive error covariances estimation methods for ensemble Kalman filters.