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

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

To cite this version:

Pierre Tandeo, Pierre Ailliot, Marc Bocquet, Alberto Carrassi, Takemasa Miyoshi, et al.. Joint Estimation of Model and Observation Error Covariance Matrices in Data Assimilation: a Review. 2018. hal-01867958

HAL Id: hal-01867958
https://hal-imt-atlantique.archives-ouvertes.fr/hal-01867958

Submitted on 4 Sep 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Joint Estimation of Model and Observation Error Covariance Matrices in Data Assimilation: a Review

PIERRE TANDEO*
IMT Atlantique, Lab-STICC, UBL, Brest, France & RIKEN Advanced Institute for Computational Science, Kobe, Japan

PIERRE AILLIOT
Laboratoire de Mathématiques de Bretagne Atlantique, University of Western Brittany, Brest, France

MARC BOCQUET
CEREA Joint Laboratory École des Ponts ParisTech and EDF R&D, Champs-sur-Marne, France

ALBERTO CARRASSI
Nansen Environmental and Remote Sensing Center, Bergen, Norway

TAKEMASA MIYOSHI
RIKEN Advanced Institute for Computational Science, Kobe, Japan

MANUEL PULIDO
Universidad Nacional del Nordeste and CONICET, Corrientes, Argentina

YICUN ZHEN
IMT Atlantique, Lab-STICC, UBL, Brest, France

ABSTRACT

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.

1. Introduction

Data Assimilation (hereinafter denoted DA) for geosciences is generally formulated in terms of nonlinear state-space models with additive and Gaussian errors for the dynamical and observation equations. This is statistically convenient and representative of a lot of DA problems, see e.g. Carrassi et al. (2018). The errors on the dynamics and observations are assumed to be zero-mean Gaussian vectors with covariance matrices $Q$ and $R$. Using the discrete time index $k$ from 1 to $K$ for the sake of simplicity, it is assumed that

$$x(k) = \mathcal{M}(k-1, x(k-1)) + \eta(k),$$

$$y(k) = \mathcal{H}(k, x(k)) + \epsilon(k),$$

with $\mathcal{M}$ the time dependent dynamical model, $\mathcal{H}$ the time dependent transformation operator from the hidden state...
x to the noisy observations y, Gaussian white noise errors
\( \eta(k) \sim \mathcal{N}(0, Q(k)) \) and 
\( \varepsilon(k) \sim \mathcal{N}(0, R(k)) \). We suppose that
the initial state condition at \( k = 0 \) is a Gaussian vector
with mean \( x^0 \) and variance \( B \) and that \( \eta \) and \( \varepsilon \) are mutually
independent. However, in some situations it may be
relevant to consider cross-correlation between these errors
(see Berry and Sauer (2018) for more details).

DA algorithms are used to estimate sequentially the
state of the system x conditionally to the observations y.
When current and past observations are used, the estimation
is referred as filtering or analysis and when future
observations are also used, it is referred as smoothing or
reanalysis. The outcome of the analysis/filter or of the re-
analysis/smoothers highly depends on the uncertainty asso-
ciated to observations and to the model state, which have
to be as realistic as possible. Using the formulation given
in Eqs. (1-2) these uncertainties are \( Q \) and \( R \). In practice,
the observation error covariance matrix \( R \) in Eq. (2)
can be determined empirically by estimating the instru-
ment noise and the representativeness error between the
state and the observation space, but a correct estimation of
the latter is often challenging (see Janjić et al. (2017)
for more details). During the dynamical model evolution
from \( k = 1 \) to \( k \) in Eq. (1), the model state is contami-
nated by two sources of uncertainty, the error in the state
at \( k = 1 \) and the model uncertainty itself which is repre-
sented in state-space models via the additive model error
term \( \eta \). Determining the covariance matrix \( Q \) model error
is difficult because it accounts for the model deficien-
cies to represent the underlying physics, the cumulative
effects of errors in the parameters, the numerical schemes,
the unresolved scales and the fact that in geosciences, we
usually have much less observations that those needed to
estimate the entries of \( Q \) (see e.g. Daley (1992) and Dee
(1995)). When using either variational or ensemble-based
DA methods, the quality of the reconstructed state vec-
tor highly depends on the relative amplitudes between the
assumed observation and model errors. For instance, in
Kalman filter-like, the ratio \( ||Q||/||R|| \) impacts the filter
gain that gives the relative weights of the observations
against the model forecasts. Desroziers and Ivanov (2001)
also studied this ratio in variational DA. Unfortunately, in
real DA frameworks, the impact of \( Q \), \( R \) and \( ||Q||/||R|| \)
on the reconstruction of the state is not easy to evaluate.
This is due to the complexity and size of the dynamical
models, the effect of forcing terms and the huge variety of
observations.

The importance of estimating error covariance matrices
in state-space models can be illustrated using a simple ex-
ample with linear dynamics. Suppose that we aim at track-
ing a scalar state \( x \) governed by an autoregressive AR(1)
model in Eq. (1) defined by
\[
x(k) = 0.95x(k-1) + \eta(k),
\]  
with \( \eta \sim \mathcal{N}(0, Q^r) \) where the superscript \( r \) means “true”
and \( Q^r = 1 \). Furthermore, observations \( y \) of the state
are contaminated with another independent additive zero-
mean and unit-variance Gaussian noise (i.e. \( R^f = 1 \)) in
Eq. (2) with \( \mathcal{N}(x) = x \). The goal is to reconstruct \( x \) from
the noisy observations \( y \) at each time step. The AR(1)
model defined by Eq. (3) has an autoregressive coeffi-
cient close to one and thus represents a process which
evolves slowly in time. The linear dynamical model
evolves stochastically and the measurement process also
introduces a noise at each time step. Although, the knowl-
edge of these two sources of noise is crucial for the es-
timation problem, in practice identifying them is not an
easy task. Given that the dynamical model is linear and
the error terms are additive and Gaussian in this simple
example, the Kalman smoother provides an exact algo-
rithm to compute the smoothing distribution (see Sect. 2
for more details). To evaluate the impact of badly specified
\( Q \) and \( R \) errors on the reconstructed state with the Kalman
smoother, different experiments were conducted using values
of \{0.1, 1, 10\} for the ratio \( Q/R \). Figure 1 shows, as a
function of time, the true state (red line) and the smoothing
Gaussian distributions represented by the 95% confidence
intervals (gray shaded) and their means (black line). We
also report the Root Mean Squared Error (RMSE) of the
reconstruction as well as the so called “coverage probabil-
ity” or percentage of \( x \) falling in the 95% confidence in-
tervals (defined as the mean \( \pm 1.96 \) the standard deviation
in the Gaussian case). In this synthetic experiment, the
best RMSE and coverage probability obtained using the
Kalman smoother with true \( Q^r = R^f = 1 \) are respectively
0.71 and 95%. When using a low model error variance
\( Q = 0.1Q^r \) in Fig. 1(a), it gives an important weight to
the forecast given by quasi-persistent autoregressive dy-
amical model. On the other hand, when using a low ob-
servation error variance \( R = 0.1R^f \) in Fig. 1(b), too much
weight is given to the observation and the reconstructed
state is too close to the noisy measurements. These results
show the negative impact of independently badly scaled
\( Q \) and \( R \) error variances. In the case of overestimated model
error variance as in Fig. 1(c), the mean reconstructed state
vector and so its RMSE are similar to Fig. 1(b). In the
same way, overestimated observation error variance as in
Fig. 1(d) gives similar mean reconstruction than Fig. 1(a).
These two last results are due to the fact that in both cases,
the ratio \( Q/R \) are equal, respectively to 10 and 0.1. Now,
we consider in Fig. 1(e) and Fig. 1(f) the case where \( Q/R \)
ratio is good (equal to 1), but respectively using the simul-
taneous underestimation and overestimation of model and
observation errors. In both cases, the mean reconstructed
state is equal to the one obtained with the true error vari-
ances, i.e. \( \text{RMSE}=0.71 \). The main difference is the gray
confidence interval which is supposed to contain 95% of
the true trajectory: the spread is clearly underestimated.
in Fig. 1(e) and overestimated in Fig. 1(f) with respective coverage probability of 36% and 100%.

Finally, let us consider again the Kalman smoother where one of the variances $Q$ or $R$ is erroneous. This time, we compensate the error in the wrong prescribed variance by optimizing the other free variance using the maximum likelihood estimation method given in Shumway and Stoffer (1982), see Sect. 4. Results presented in Fig. 2 show that best optimal RMSE (0.71) and coverage probability (95%) are reached close to the optimal variance noises $Q^* = R^* = 1$. Results also indicate that a compensation of bad variances is possible but is not optimal. For instance, when fixing $Q$ to a bad value like $0.25$ in Fig. 2(a) and Fig. 2(b), the maximum likelihood estimator of $R$ is $1.57$ and corresponding RMSE and coverage probability are respectively $0.86$ and $80\%$ (out of range of the color bar). These two skill metrics are extremely important to evaluate the quality of the reconstructed state. Nevertheless, only the RMSE is presented in research papers. The coverage probability is a measure of the DA capability to quantify the uncertainty, a problem that we believe to be of increasing relevance for the DA community in the coming years. Indeed, the reconstructed state error variance may have a strong impact and may make filters to collapse. In this linear and Gaussian example, the use of RMSE and the probability of coverage are sufficient, but for nonlinear and more realistic DA cases, we should also consider the rank histograms and the proper scores.

Since the 90’s, a significant number of works have dealt with the error covariances in state-space models. The first ones who mentioned the importance of noise covariance matrices $Q$ and $R$ in DA were Ghil and Malanotte-Rizzoli (1991) in their Sect. 4.1, as well as Daley (1991) in his Sect. 4.9. Daley (1992) clarified the difference between “predictability error” and “model error”, the two components of the forecast error, denoted as $P^f$ in modern DA. As illustrated in Fig. 3, the first error is due to imperfect initial conditions and the second one is caused by model imperfections represented by $Q$. Dee (1995) proposed a maximum likelihood estimator for parameterized versions of $Q$ and $R$ using the innovation likelihood criterion. Dee et al. (1999a) extended this online method to the estimation of the mean of the innovations, which depends on the biases in the forecast and in the observations, and later applied to realistic cases in Dee et al. (1999b). These initial studies clearly impulsed the treatment of this topic in modern DA literature and several works have appeared thereafter on the joint estimation of model and observation errors. However, authors like Todling (2015) pointed out that using only the current innovation is not enough to dis-
Figure 2. Estimation results of $R$ and $Q$ for respectively fixed $Q$ and $R$ in the case of a Kalman smoother with the univariate AR(1) process given in Eq. (3). Here, we consider the maximum likelihood estimator given by Shumway and Stoffer (1982).

tinguish the impact of $Q$ and $R$ in the Kalman equations, and to estimate them independently is challenging in this case. Thus, they proposed various alternatives to tackle this issue.

An history of what have been, in our opinion, the most relevant contributions and the key milestones for covariance estimation in geophysical systems is sketched in Fig. 4 and is discussed in this review with a summary given in Table 1. We distinguish four methodologies and among them, one could classify the approaches whether they rely upon the innovations or the likelihood. The innovations are defined as the difference between the observations and state estimates transformed to the observational space, with both the forecast and the analysis. The use of their corresponding statistics in the observation space has been initiated by Desroziers et al. (2005). Then, this approach has been used extensively for the calibration of inflation of the forecast covariance with various implementations including additive, relaxation-to-prior and the multiplicative inflation case, see Li et al. (2009a) and Miyoshi (2011). Instead of working on different innovations at a given time, Berry and Sauer (2013) as well as Harlim et al. (2014) suggested to use lag-innovations or innovation between consecutive times. At the same time, methods based on likelihood functions and their maximization using statistical approaches appeared. Bayesian inference techniques with the use of prior distributions and hyperparameters as in Stroud and Bengtsson (2007) or Stroud et al. (2018) are typical examples. Finally, Ueno and Nakamura (2014), Dreano et al. (2017) and Pulido et al. (2018) proposed to maximize the total likelihood of the state-space model using iterative expectation-maximization algorithms.

The four methods mentioned above are detailed in this review and are factually described as they appear in the literature. We consider both online and offline estimations, for which the computational cost highly varies. In the online or adaptive approaches, we try to estimate a time-dependent $Q(k)$ and $R(k)$ at the same time as the state vector, using filtering methods. When considering offline or batch approaches, averaged $Q$ and $R$ are estimated using all the observations on a given time interval, using smoothing methods. Moreover, offline procedures are iterative, meaning that the procedures are repeated until convergence according to a given criterion, for instance the likelihood. Finally, in some methods presented here, additional tuning parameters are needed and have to be carefully chosen for practical implementations. We discuss this point in this review.

Note that other review papers on parameters estimation in state-space models appeared in the statistical and signal processing communities by Mehra (1972), Kantas et al. (2015) and Dunik et al. (2017). The Mehra (1972) paper is a concise review which accounts for linear dynamical models using the classic Kalman filter while Kan-
Short description of the extended version of the Kalman equations for nonlinear dynamical systems and observation operators. Here we use time dependent linearizations $M$ and $H$ of the nonlinear operators $\mathcal{M}$ and $\mathcal{H}$ defined in Eqs. (1-2). We have chosen to base the discussion on the Extended Kalman Filter and Smoother (EKF/EKS) in this review, compared to Ensemble Kalman Filter (EnKF) for instance, to avoid the overburdening notations introduced by the ensemble members. However, the methods are also straightforward to apply in stochastic and square-root EnKFs.

Note that the most natural algorithms to solve the state-space model given in Eqs. (1-2) are the Particle Filter and Smoother (PF and PS) from Gordon et al. (1993) and firstly reviewed in DA by van Leeuwen (2009). These methods converge to the true posterior distributions for a large number of particles in theory. However, we focus in this review on Gaussian additive errors $\eta$ and $\xi$ in Eqs. (1-2), and EKF/EKS perform generally well in this situation. Moreover, the current PF and PS implementations are subject to the curse of dimensionality (Snyder et al. (2008)) and are not suitable for high dimensional systems, although recent implementations appear to shed some light on these contentious points (e.g., Atkins et al. (2013) and Zhu et al. (2016)). However, because of PF relies on a state-space model formulation as in Eqs. (1-2), the model error covariance specification is an essential requirement in the definition of the transition density. Therefore, the estimation of $Q$ with Gaussian Kalman-based methods may give a useful constraint or parameterization setup of the model error covariance matrix for PF (see Zhu et al. (2017)).

Kalman-based algorithms assume a Gaussian prior distribution $p(\mathbf{x}(k)|y(1:k-1)) \sim \mathcal{N}(\mathbf{x}^f(k), \mathbf{P}^f(k))$. Then, filtering and smoothing estimates are corresponding to the Gaussian posterior distributions $p(\mathbf{x}(k)|y(1:k)) \sim \mathcal{N}(\mathbf{x}^a(k), \mathbf{P}^a(k))$ and $p(\mathbf{x}(k)|y(1:K)) \sim \mathcal{N}(\mathbf{x}^a(k), \mathbf{P}^a(k))$ of the state conditionally to past/present observations and past/present/future observations respectively. Here, we briefly remind the equations of the EKF and EKS based on the Rauch-Tung-Striebel (RTS) solution detailed in Cosme et al. (2012). They are divided in three main steps:

2. Filtering and smoothing algorithms

For the overall discussion of the methods and for introduction of the notation, we present in this section a
Forecast step (forward in time):

\[ x^f(k) = \mathcal{M}(x^f(k-1)) \]  \hspace{1cm} (4)

\[ P^f(k) = M(k)P^f(k-1)M^\top(k) + Q(k) \]  \hspace{1cm} (5)

Analysis step (forward in time):

\[ d(k) = y(k) - \mathcal{H}(x^f(k)) \]  \hspace{1cm} (6)

\[ K^f(k) = P^f(k)H(k)^\top \left( H(k)P^f(k)H(k)^\top + R(k) \right)^{-1} \] \hspace{1cm} (7)

\[ x^a(k) = x^f(k) + K^f(k)d(k) \] \hspace{1cm} (8)

\[ P^a(k) = (1 - K^f(k)H(k))P^f(k) \] \hspace{1cm} (9)

Reanalysis step (backward in time):

\[ K^s(k) = P^a(k)M(k)^\top \left( P^f(k+1) \right)^{-1} \] \hspace{1cm} (10)

\[ x^s(k) = x^a(k) + K^s(k)(x^f(k+1) - x^f(k+1)) \] \hspace{1cm} (11)

\[ P^s(k) = P^a(k) \] \hspace{1cm} (12)

\[ P^s(k, k+1) = P^s(k+1)K^s(k)^\top \] \hspace{1cm} (13)

3. Innovation-based methods

The importance of the innovation statistics has been emphasized in the DA community by Daley (1992) and Dee (1995). The “classic innovation” \( d \), difference between the observations and the forecasted states in the observation space, defined in Eq. (6), implicitly takes into account the \( Q \) and \( R \) covariances. Unfortunately, as explained in Blanchet et al. (1997), by using only current observations, their individual contributions cannot be easily disentangled. Thus, the approaches using only the classic innovations are not studied in this review. Two main approaches were proposed in the literature to tackle this issue. They are based on the idea of producing multiple equations involving \( Q \) and \( R \). The first one uses different innovation statistics in the observation space. The second one is based on lag-innovations or differences between consecutive innovations. From a statistical point of view, the innovation-based methods are “methods of moments”, where we construct a system of equations which links various moments of the innovations with the parameters and then replace theoretical moments by the empirical ones in these equations.

3.1. Innovation statistics in the observation space

Desroziers et al. (2005) proposed to examine various innovation statistics in the observation space. This method is now popular in the DA community. It is based on different innovation statistics between observations, forecasts and analysis, and all of them defined in the observation space: namely, \( d^{o-f}(k) = y(k) - \mathcal{H}(x^f(k)) \) as in Eq. (6) and \( d^{o-a}(k) = y(k) - \mathcal{H}(x^a(k)) \). We remark that another diagnostic using the difference between analysis \( x^a(k) \) and reanalysis \( x^s(k) \) has been proposed by Todling (2015) and
Bowler (2017) in the case of sequential and variational DA respectively to estimate the covariance $Q$ alone. In theory, in the linear and Gaussian case, the Desroziers innovation statistics should verify the equalities:

$$
\begin{align*}
E \left[ d^{o-f}(k) d^{o-f}(k)^\top \right] &= H(k)P^f(k)H(k)^\top + R(k) \quad (14) \\
E \left[ d^{o-a}(k) d^{o-a}(k)^\top \right] &= R(k) 
\end{align*}
$$

with $E$ the expectation operator. In this approach, we do not estimate $Q$ directly which is implicitly taken into account in $P^f$. Instead, the approach attempts to compensate in $P^f$ for the lack of knowledge of $Q$ as well as the systematic variance underestimation. This method is referred to as “covariance inflation”. In practice, when using for instance EnKF with a small ensemble size, the spread is most of the time underestimated and this leads to filter divergence (see e.g. Carrassi et al. (2018), their appendix A). Thus, covariance inflation can be required in a perfect model scenario (i.e. without $\eta$), because of sampling errors. For imperfect models, both sampling errors and an inappropriate representation of model errors lead to an underestimation of forecast ensemble spread and thus to filter divergence, see Raanes et al. (2018).

We distinguish three inflation methods: multiplicative, additive and relaxation-to-prior. In the multiplicative case, the forecast error covariance matrix $P^f$ is usually multiplied by a scalar coefficient greater than 1, see Anderson and Anderson (1999). Adaptive procedures to estimate this coefficient have been proposed by Wang and Bishop (2003), Li et al. (2009a), Miyoshi (2011) and Bocquet (2011) in the case of innovation statistics in the observation space. In the additive case, the diagonal of the forecast and/or analysis empirical covariance matrices is increased (Mitchell and Houtekamer (2000), Corazza et al. (2003), Whitaker et al. (2008) and Houtekamer et al. (2009)). In the relaxation-to-prior case, Zhang et al. (2004) blended the forecast and analysis ensemble perturbations whereas Whitaker and Hamill (2012) multiplied the analysis ensemble spread to relax the reduction of the spread, without blending perturbations. Finally, Bocquet and Sakov (2012), Ying and Zhang (2015) and Kotsuki et al. (2017) proposed methods to adaptively estimate the relaxation parameters using innovation statistics. Adaptive covariance inflations are online estimation methods directly plugged to classic filtering method (like EKF here), with almost no additional computational cost. In practice, the use of this technique does not necessarily imply an additive error term $\eta$ in Eq. (1). Thus, it is not a direct estimation of $Q$ but an inflation applied to $P^f$ in order to compensate model uncertainties and sampling errors in EnKFs (see Raanes et al. (2018), Sect. 4 and appendix C for more details). Several DA systems work with an inflation method and used it for its simplicity, low-cost and efficiency.

Here, we focus on the straightforward online estimation of a multiplicative inflation factor $\lambda$ of the badly scaled $P^f(k)$ so that the corrected forecasted covariance is given by $P^f(k) = \lambda(k) P^f(k)$. The estimate of the inflation factor is given by taking the trace of Eq. (14):

$$
\lambda(k) = E \left[ \frac{d^{o-f}(k) d^{o-f}(k) - Tr(R(k))}{Tr(H(k)P^f(k)H(k)^\top)} \right]. 
$$

The use of temporal smoothing for the online estimation of $\lambda(k)$ is crucial in operational procedures and Miyoshi (2011) proposed augmenting the state vector with the inflation factor whose evolution is governed by a random walk equation. In this case, we need to specify an additional parameter for the variance term of this random walk, denoted by $\sigma^2$. This parameter has to be carefully tuned to avoid the divergence of the filter. Then, at each time step $k$, when sufficient observations are available, an estimate of $R(k)$ is directly given by Eq. (15). Note that Li et al. (2009a) proposed to estimate each component of a diagonal $R$ matrix, and also suggested to use an offline procedure to compute the average of these variance terms.

### h. Lag-innovation between consecutive times

Another way to estimate error covariances is to use multiple equations involving $Q$ and $R$ exploiting cross-correlations between lag-innovations, i.e. the current $d(k)$ and past classic innovations $d(k-1)$, $\ldots$, $d(k-l)$. For instance, considering the lag-zero and lag-one innovations, the following equations are satisfied in the linear and Gaussian case:

$$
\begin{align*}
E \left[ d(k) d(k)^\top \right] &= H(k)P^f(k)H(k)^\top + R(k) = \Sigma(k) \\
E \left[ d(k) d(k-1)^\top \right] &= H(k)M(k)P^f(k-1)H(k)^\top - H(k)M(k)n^f(k-1)\Sigma(k-1). 
\end{align*}
$$

Lag-innovations were introduced by Mehrha (1970) in order to simultaneously recover the error covariance matrices for a Gaussian and linear state-space model. Mehrha established analytic exact relations between $Q$ and $R$, and the probabilistic expectations of $d(k)d(k-l)^\top$ for linear systems in steady state. Then, Bélanger (1974) extended these results to the case of time-varying linear stochastic processes, taking $d(k)d(k-l)^\top$ as “observations” of $Q$ and $R$ and using a secondary Kalman filter to update them iteratively. As pointed out in Bélanger (1974), this method would no longer be analytically exact if the error matrices are updated adaptively at each time step. Later, Dee et al. (1985) proposed a computationally cheaper algorithm for the Bélanger’s method. More recently, authors focused on high-dimensional and nonlinear systems using the EKF and EnKF: Berry and Sauer (2013) proposed a fast algorithm based on Mehrha’s method and Harlim
et al. (2014) followed the original Bélanger’s algorithm. Zhen and Harlim (2015) proposed a modified version of Bélanger’s method and compared it to the Berry and Sauer (2013) approach.

Here, we briefly introduce the algorithm of Berry and Sauer (2013) using lag-one innovations. It is based on the online (or adaptive) estimation of $Q(k)$ and $R(k)$, which satisfy the following relations in the linear and Gaussian case:

$$
\dot{P}(k) = M(k)^{-1}H(k)^{-1}d(k)d(k-1)^{\top}H(k)^{-T} + K/\tau(k-1)d(k-1)^{\top}H(k)^{-T},
$$

(19)

$$
\dot{Q}(k) = \dot{P}(k) - M(k-1)P(k-1)^{-T}M(k-1)^{\top},
$$

(20)

$$
\dot{R}(k) = d(k-1)d(k-1)^{\top} - H(k)P(k-1)^{-T}H(k)^{\top},
$$

(21)

In this online procedure, joint estimations of $\dot{Q}(k)$ and $\dot{R}(k)$ can abruptly vary over time. Thus, the temporal smoothing of the covariances being estimated becomes crucial. As suggested by Berry and Sauer (2013), an exponential smoothing between current and past estimates is a reasonable choice,

$$
Q(k+1) = Q(k) + (\dot{Q}(k) - Q(k)) / \tau,
$$

(22)

$$
R(k+1) = R(k) + (\dot{R}(k) - R(k)) / \tau,
$$

(23)

with $\tau$ the smoothing parameter and started from $Q(0)$. When $\tau$ is small, weight is given to the current estimate $Q$ and when $\tau$ is larger it gives a smoother sequence $Q$. As pointed out by Zhen and Harlim (2015), usually a large value of $\tau$ is chosen to avoid numerical instability.

It is worth pointing out that in the case of sparse observations, the estimate of $\dot{P}$ in Eq. (19) might be underdetermined, even if the system is observable. This is attributed to the use of only one lag-innovation. Theoretically, all components of $Q$ should be identifiable if the system is observable and more lag-innovations are used. But in practice, using more lag-innovations implies increased computational cost and does not necessarily lead to accurate estimates. Zhen and Harlim (2015) compared the modified version of Bélanger’s method with different choices of maximal lags and found that a maximal lag of 4 is optimal in a specific numerical example on Lorenz-96 defined in Lorenz (1996).

4. Likelihood-based methods

The likelihood approaches were put forward in the DA community by Dee (1995), Blanchet et al. (1997) as well as Mitchell and Houtekamer (2000) where it was proposed to maximize the likelihood of the innovation, i.e. $p(y(k)|y(k-1))$, defined by the mean vector $d(k)$ computed in Eq. (6) and covariance matrix $X(k)$ introduced in Eq. (17) and also used in the computation of the Kalman filter gain in Eq. (7). Unfortunately, they reach the same conclusions than for the innovation-based methods, i.e. the joint estimation of $Q$ and $R$ is not straightforward if we use only the current observations. To tackle this issue, several methods have been proposed recently. The first one is to write the estimation problem using a Bayesian framework, and jointly estimate prior distributions of $Q$ and $R$ parameters with the innovation likelihood. The second one is to maximize the so-called “total likelihoods”, i.e. taking into account the innovation likelihoods of each time step or taking into account the global structure of the state-space model for all the time steps.

a. Bayesian inference

In a Bayesian approach, we assume that the elements of $Q$ and $R$ covariance matrices have a priori distributions which are controlled by some hyperparameters. In practice, it is difficult to have a prior distribution for each element of $Q$ and $R$, especially for large DA systems. Instead, parametric forms are used for the matrices, typically describing the shape and level noise, and we denote the corresponding parameters as $\theta$. Then, we jointly and adaptively estimate the state $x$ and parameters $\theta$ using Bayes’ theorem:

$$
p(x(k), \theta(k)|y(1:k)) = p(x(k)|y(1:k), \theta(k)) p(\theta(k)|y(1:k)).
$$

(24)

In Eq. (24), $p(x(k)|y(1:k), \theta(k))$ is given by filtering DA algorithms and we approximate recursively $p(\theta(k)|y(1:k))$ using the the likelihood of the innovations $p(y(k)|y(1:k-1), \theta(k))$ as

$$
p(\theta(k)|y(1:k)) \propto p(y(k)|y(1:k-1), \theta(k)) p(\theta(k)|y(1:k-1)).
$$

(25)

Bayesian approaches have been applied in the atmospheric chemistry community and reviewed by Michalak et al. (2005) and Wu et al. (2013). Purser and Parrish (2003) introduced the Bayesian approach in variational DA for the estimation of two statistical parameters, controlling the magnitude of the variance and the spatial dependencies in $Q$, assuming that $R$ is known and using a univariate model. Then, Stroud and Bengtsson (2007) used a similar approach combined with EnKF in the Lorenz-96 model for the estimation of a common multiplicative scalar parameter for predefined matrices $Q$ and $R$. In that case, the scalar parameter affects simultaneously the $Q$ and $R$ matrices. Considering the experiments about the importance of $||Q||/||R||$ ratio presented in Fig. 1, we can guess that this approach is maybe not optimal. Then, other works have applied similar Bayesian approaches for the estimation of parameters governing the shape of $R$ only: Frei and Künisch (2012) in the Lorenz-96 system, Winiarek et al. (2012, 2014) assimilating nuclear pollutants using a regional atmospheric model (in
this case, $\mathbf{R}$ partially accounts for model error), Ueno and Nakamura (2016) using two linear shallow-water equations to assimilate satellite altimetry. By contrast, Solonen et al. (2014) proposed a Bayesian approach for the estimation of $\mathbf{Q}$ only, assuming that the $\mathbf{R}$ matrix is known, in a two-layer quasi-geostrophic model. Finally, Stroud et al. (2018) tested their estimation method on different spatio-temporal systems with a joint estimation of $\mathbf{Q}$ and $\mathbf{R}$. Their experiments on the Lorenz-96 system assumed no model error $\mathbf{Q}$.

The Bayesian inference approach is an online estimation procedure with joint estimation of the state of the system and hyperparameters $\theta(k)$, controlling shape parameters of the $\mathbf{Q}(k)$ and $\mathbf{R}(k)$ error covariance matrices. In terms of the hidden state, this corresponds to a hierarchical Bayesian approach so that an ensemble of filters may be required to determine the posterior distribution in Eq. (24). In practice, this method may required a large number of Monte-Carlo simulations to estimate correctly $p(\theta(k)|y(1:k))$ defined in Eq. (25) or iterative procedures as in Ueno and Nakamura (2016). Alternatively, Scheffler et al. (2018) assume a Gaussian distribution for $p(\theta(k)|y(1:k-1))$ in Eq. (25) and use two nested EnKFs, reducing the computational cost of this hierarchical Bayesian procedure. In principle, the Bayesian approach is able to estimate time dependent hyperparameters, however recent works assume that $\theta(k)$ is constant, not depending in time. The potential of this framework to estimate time dependent covariances is an interesting topic to be addressed. The joint estimation of parameters controlling separately $\mathbf{Q}$ and $\mathbf{R}$ still remains a challenge. However relevant parametric shapes of covariance matrices, such as the Matérn covariance model for $\mathbf{R}$, have been proposed as in Stroud et al. (2018).

b. Maximization of the total likelihoods

The innovation likelihood at time $k$ is defined by $p(y(k)|y(1:k-1), \theta(k))$ in Eq. (25). Maximizing this likelihood at each time step has been proposed by various authors: Dee (1995), Blachet et al. (1997), Zheng (2009), Mitchell and Houtekamer (2000) and Liang et al. (2012). But the maximization of the innovation likelihood has two main issues. Firstly, the innovation covariance matrix $\Sigma(k) = \mathbf{H}(k)\mathbf{P}^{-1}(k)\mathbf{H}(k)^\top + \mathbf{R}(k)$ is mixing the information of $\mathbf{R}$ and $\mathbf{Q}$, contained in $\mathbf{P}^{-1}$. When using only one $k$, it is difficult to identify the model and observation error covariances and in practice, authors only estimated one of them. Secondly, the number of observations at each time step is in general limited and as pointed out by Dee (1995), available observations should exceed “the number of tunable parameter by two or three orders of magnitude”. For these reasons, a reasonable alternative is to use a batch of observations distributed in time and assume $\theta$ being constant in time. The resulting total likelihood expressed sequentially through conditioning is given by

$$ p(y(1:K)|\theta) = \prod_{k=2}^{K} p(y(k)|y(1:k-1), \theta). \quad (26) $$

This likelihood is said to be “incomplete” because it only depends on the observations of the state-space system, not the hidden state. But since this is an integration innovation likelihood over a long period of time, this gives more information to estimate $\mathbf{Q}$ and $\mathbf{R}$ or related parameters. This likelihood is said to be marginal since it results from marginalizing the hidden state at a given observation time. The incomplete total likelihood is also a useful tool to evaluate the quality of model forecasts or how well they match the observations, considering both model and observation uncertainties. Hannart et al. (2016) and Carrassi et al. (2017) used it for model evidence, when various models are in competition. The maximization of the incomplete total likelihood given in Eq. (26) has been applied to linear and nonlinear systems in the estimation of deterministic and stochastic parameters (related to $\mathbf{Q}$) in Delsole and Yang (2010) using a direct sequential optimization procedure. Then, for nonlinear dynamics, Ueno et al. (2010) used a grid-based procedure to estimate noise levels and spatial correlation lengths of $\mathbf{Q}$ and a level noise for $\mathbf{R}$. This grid-based method used predefined sets of covariance parameters and test the different combinations to find the one that maximizes the likelihood criterion. Brankart et al. (2010) proposed also a method using the same criterion additionally with an initial information on scale and correlation length parameters of $\mathbf{Q}$ and $\mathbf{R}$. This information is only given at the first time and progressively forgotten with time using a decreasing exponential factor. The marginalization of the hidden state in Eq. (26) is considering all the previous observations and in practice it requires the use of a filter. The maximization of the total incomplete likelihood using the EnKF to estimate model error covariance $\mathbf{Q}$ was conducted in Pulido et al. (2018) where they used a gradient-based optimization technique.

Authors also proposed to work on the maximization of the total likelihood using the marginalization of the whole trajectory of the hidden state from $k = 0$ to $K$. In that case, we talk about the “complete” total likelihood or joint density of the observations and the hidden state, expressed as

$$ p(y(1:K), x(0:K)|\theta) = \prod_{k=1}^{K} p(x(k)|x(k-1), \theta) \prod_{k=1}^{K} p(y(k)|x(k), \theta) $$

where the three terms on the rhs are related to the initial state, the state equation in Eq. (1) and the observation equation in Eq. (2). In practice, the marginalization of the full hidden state from $k = 0$ to $K$ is not possible,
so that the complete total likelihood cannot be evaluated

directly, see explanations in Pulido et al. (2018). Therefore,
Shumway and Stoffer (1982) proposed to use an iterative
procedure, requiring the use of a smoother, to max-
imize the likelihood criterion given in Eq. (27). They
used the Expectation-Maximization algorithm (hereinafter
noted EM, see Dempster et al. (1977) for more details) to
ensure the convergence to the maximum likelihood es-
timator, and applied it to estimate \( Q \) and \( R \) in the case of
linear dynamics. In DA, the EM algorithm has been
implemented for estimating only \( R \) in Ueno and Nakamura
(2014) for the Zebiski and Cane (ZC) model and satellite
altimetry, for the background covariance matrix \( B \) and \( R \) in
Liu et al. (2017) in case of accidental pollutant source re-
trieval, for both \( Q \) and \( R \) with an orographic subgrid-scale
nonlinear observation operator in Tandeo et al. (2015),
and for the Lorenz-63 in Dreano et al. (2017). Recently,
Pulido et al. (2018) used the EM algorithm and compared
with a gradient ascent optimization of the incomplete total
likelihood to estimate \( Q \) along with the deterministic and
stochastic physical parameters in the one and two scales
Lorenz model described in Lorenz and Emanuel (1998).
Finally, Chau et al. (2018) combined conditional particle
filters and the EM algorithm for the joint estimation of
\( Q \) and \( R \) and show the improvement compared to Kalman-
based filters.

The EM algorithm considers the total period of time
\( k = 0 \) to \( K \) to maximize complete the total likelihood given
in Eq. (27). Thus, it leads to an offline estimation of the
error covariance matrices or related parameters (i.e. con-
stant, non-adaptive). In the expectation step, we value
the expected likelihood of the previous estimates of \( Q \) and \( R \) as well as the to-
tal observations \( \{y(1), \ldots, y(K)\} \). This leads to the use of
Kalman smoother procedures to estimate \( \hat{x}^{k} \) and \( P^k \) at the different times.
In the maximization step, \( Q \) and \( R \) are
updated using the following estimators:

\[
Q = \frac{1}{K} \sum_{k=1}^{K} \{ (x^{k}(k) - \mathcal{H}(x^{k}(k - 1))) (x^{k}(k) - \mathcal{H}(x^{k}(k - 1)))^\top + M(k)P^k(k - 1)M(k)^\top + P^k(k) \}
- P^k(k - 1, k)M(k)^\top - M(k)P^k(k - 1, k)^\top \} \tag{28}
\]

\[
R = \frac{1}{K} \sum_{k=1}^{K} \{ (y(k) - H(k)x^{k}(k)) (y(k) - H(k)x^{k}(k))^\top + H(k)P^k(k)H(k)^\top \} \tag{29}
\]

In practice, the total and complete likelihood given in
Eq. (27) cannot be evaluated exactly. Thus, to evaluate
the performance at each iteration of the EM procedure, we
compute the incomplete total likelihood given in Eq. (26)
and this criterion increases along the iterations of the EM
algorithm, see Wu (1983). The computational cost of
the EM algorithm is large because it requires the use of
Kalman-based filter and smoother at each iteration. Nev-
evertheless, this EM method does not require any additional
parameter and is robust to initial conditions, i.e. the val-
ues given for \( Q \) and \( R \) matrices in the first EM iteration.
Additionally, note that the maximization of the total likely-
hood allows the estimation of the initial state vector \( x^0 \)
and covariance matrix \( B \) as discussed in Tandeo et al. (2015)
and Dreano et al. (2017). In that case, we should write
\( p(x(0)|\theta) \) instead of \( p(x(0)) \) in Eq. (27).

5. Other methods

In this section, we describe other methods which have
been used in the past or methods that are more diagnostic
tools than direct estimations of error covariance matrices.
We also include some relevant references on inverse prob-
lems in environmental data.

a. State augmentation

State augmentation was first proposed in Schmidt
(1966) and is known as the Schmidt-Kalman filter. Then,
Jazwinski (1970) proposed some extensions. The idea is to
augment the state vector in order to estimate both the state
of the system and additional parameters among which the bias,
forcing terms, physical parameters, and finally error covari-
cances as in Zupanski (1997) and Tremolet (2007).
In these works, authors create cross-correlation between
the state of the system and the additional parameters. The
method works only for parameters strongly related to the
state of the system, such as physical parameters, see Ruiz
et al. (2013). However, Stroud and Bengtsson (2007) as
well as Delsole and Yang (2010) formally demonstrated
that augmentation methods fail for variance parameters
and thus for \( Q \) and \( R \). Indeed, a critical aspect of this ap-
proach is that one needs to define an evolution model for
the covariance parameters. This is a difficult task, and often
persistence is used, which means that the estimate and
the associate error variance only change at analysis times,
and the estimated variance is thus bound to decrease in
time: it is reduced during the forecast step using a per-
sistence model. This makes the use of random walk or
inflation mandatory, or a change in the parametric error
dynamics such as in Carrassi and Vannitsem (2011a).

b. Analysis increment approach

Analysis increments refers to statistical methods that
study the relationship between two consecutive times of
a dynamical system. The use of regression methods has
been firstly proposed by Lorenz (1977) and then by Leith
(1978) to learn error statistics of dynamical models in me-
teorology (e.g. bias and covariance \( Q \) of the error \( \eta \)).
Then, this approach was first discussed in the context of
DA by Li et al. (2009b), and it was then further expanded
by Carrassi and Vannitsem (2010) in the context of variational DA with time correlated model error. The same reanalysis increment approach has been used in Carrassi and Vannitsem (2011b) to estimate model error due to unresolved scale and later applied in the context of a deterministic EnKF by Mitchell and Carrassi (2015).

c. Covariance matching

The covariance matching method has been introduced by Fu et al. (1993). It consists in matching sample covariance matrices to their theoretical expectations. It corresponds to a method of moments applied to different innovations. Thus, it is very similar to Desroziers et al. (2005) method, except that covariance matching is performed on a set of historical observations and numerical simulations, not online in a DA scheme. It has been extended by Menemenlis and Chechelnitsky (2000) to time-lagged innovations as in Bélanger (1974) and they also relaxed the assumption of independence between the true state and model simulation errors.

d. The $\chi^2$, cross-validation and whiteness of lag-innovations

These methods are not direct estimation of error covariance matrices but diagnostic tools. The first one is a statistical test that examines the variance of the normalized innovations that follow in theory a $\chi^2$ distribution with a given number of degrees of freedom. As pointed out by respectively by Michalak et al. (2005) and Rayner et al. (1999), lots of combination of Q and R errors can lead to the acceptance of this test and this method cannot guide the relative allocation of error between the 2 ones. Cross-validation is a classic bootstrap strategy to test the accuracy in statistical modeling. It is based on the repetition of validations between a learned model from a random training dataset and test on the rest of the observations, see Wu et al. (2013) for more details. The third one was pointed by Jazwinski (1970) and consists in evaluating the properties of lag-innovations. They are supposed to be white in time in case of optimal filtering, i.e. using appropriate Q and R matrices.

6. Summary, conclusions and perspectives

In this review paper, we presented different methods to estimate the error covariances in data assimilation. As usually stated in data assimilation, we assume that model and observation errors are additive and Gaussian with centered null mean and covariance matrices noted Q and R. The individual and joint impacts of badly calibrated covariances were firstly presented using a linear toy-model. The experiments clearly showed that for reaching reasonable results of the filter, in terms of pure reconstruction using root mean squared error, the joint estimation of both components is a crucial point. We also highlighted the impact on the coverage probability, related to the estimated covariance of the reconstructed state and thus to the uncertainty quantification in data assimilation.

Summary of existing methods

We focused on four main methodologies for the joint estimation of Q and R used in data assimilation. They are summarized in Table 1. We first dealt with methods based on innovations, i.e. the difference between observations and forecasted state, and the use of empirical and theoretical moments, also known as moments of moments in statistics. We presented the method of innovation statistics in the observation space by Desroziers et al. (2005). This approach is often associated with inflation methods where model error covariance Q is not necessary needed and where an estimated inflation factor artificially increases the forecast error covariance $P'$. This online method is low-cost and adaptive but the inflation factor controls the global amplitude of the covariance and cannot adapt to specific parts of it. Nevertheless, the covariance inflation is used in lot of operational data assimilation systems and is very common with lot of implementations like the multiplicative one, see Li et al. (2009a) and Miyoshi (2011) for instance. Another approach received recently specific attention in the data assimilation context: the lag-innovation exploiting the autocorrelation of the innovation between consecutive times. It has been introduced by Mehra (1970) and Berry and Sauer (2013) extended this lag-one innovation method to the nonlinear case in data assimilation, whereas Harlim et al. (2014) implemented a lag-l innovation method following Bélanger (1974) idea. These lag-innovation techniques are adaptive, online and plugged into any filter with a moderate additional computational cost. However, results are very sensitive to a tuning parameter $\tau$, used to smooth the estimated covariance matrices along time and avoid the method from breaking down. Lag-innovation methods have been tested on linear and Lorenz-96 systems, not yet on real data assimilation schemes. At this stage, authors considered constant Q and R but in practice, when $\tau$ parameter is correctly tuned, lag-innovations can deal with time varying matrices.

The two last methods summarized in Table 1 are based on the maximization of the likelihood criterion. Dee (1995) firstly pointed out the importance of maximizing the innovation likelihood, but the estimation of both Q and R using only current observations is limited. Thus, authors then proposed Bayesian inferences to jointly maximize the innovation likelihood and the likelihood of parameters of the error covariance matrices. These parameters are assumed to follow prescribed prior distributions that have to be carefully chosen. Moreover, large Monte-Carlo simulations are needed to estimate the hyperparameters of these distributions. The joint estimation of Q and R has not been
evaluated, except in Stroud and Bengtsson (2007) where they used the same inflation parameter for both matrices. These approaches have been tested with the Lorenz-96 system in Frei and Künsch (2012) and with realistic models and observations in Winiarek et al. (2012, 2014) as well as Ueno and Nakamura (2016). Bayesian approaches are online methods and able to deal with time varying covariances. The last method is based on the maximization of the complete total likelihood as proposed in linear state-space models as shown in Dreano et al. (2017). In comparison to the previous three methods, this one is offline and not adaptive, assuming that \( Q \) and \( R \) are constant over a batch period. It has a high-computational cost but it does not require any additional tuning parameter. So far, recent works jointly estimate model and observation error matrices in toy-models like Lorenz-63 and Lorenz-96 with one and two scales, respectively in Dreano et al. (2017) and Pulido et al. (2018). Ueno and Nakamura (2014) applied the EM algorithm for the estimation of \( R \) in a realistic coupled atmosphere-ocean model and Liu et al. (2017) applied EM algorithm to a radionuclide transport model using real observations.

Remaining challenges

There are still remaining challenges for the four methods detailed in this review. The first concerns the improvements of online and adaptive techniques regarding additional parameters that control the variations of \( Q \) and \( R \) estimates in time. Instead of using fixed values for these parameters, for instance \( \tau \) in lag-innovations or \( \sigma_2^2 \) in inflation methods, we suggest to use time-dependent adaptations. This will avoid the problems of instabilities close to the solution. Another option could be to adapt these online procedures to the offline case, working with very stable parameters values (\( \tau \) high, \( \sigma_2^2 \) low) and iterate the procedures on a batch of observations as in the EM algorithm. This was suggested and rapidly tested in Desroziers et al. (2005) with encouraging results. To the best of our knowledge, it has not been yet tested with lag-innovation methods.

The second challenge concerns the Bayesian approach where joint estimation of \( Q \) and \( R \) seems problematic. As pointed out by Berry and Sauer (2018), correlation between model and observation error terms is in practice highly probable in real data assimilation problems. In this case, instead of using independent prior distributions, the use of joint prior distributions for parameters of \( Q \) and \( R \) might physically constraint the optimization procedure.

A third challenge concerns the offline EM algorithm using the total likelihood over a large batch of observations. This procedure can be adapted to account for time varying error covariances. A simple way is to work on small independent sets of observations and apply various EM procedures. Then, the \( Q \) and \( R \) could be smoothed in time. Another way is to apply online EM algorithms (see for instance Cappé (2011)) with the likelihood averaged locally in time. Note also that EM algorithm, whether for the online or offline case, can be coupled with direct optimization methods like Newton-Raphson to speedup convergence.

From our point of view, the last challenge concerns the estimation of other statistical parameters of the state-space model given in Eqs. (1-2) and associated filters. Indeed, the initial condition \( x^0 \) and \( B \) are crucial for certain satellite retrieval problems and have to be estimated, principally in offline cases where filtering and smoothing are repeated on various iterations. Finally, estimation methods should also consider the estimation of systematic or time varying biases, corresponding to the deterministic part of \( \eta \) and \( \varepsilon \). It has been initially proposed by Dee et al. (1999a) and tested in Dee et al. (1999b) in the case of the maximization of the innovation likelihood and re-

<table>
<thead>
<tr>
<th>Described method</th>
<th>Criteria</th>
<th>Estimation method</th>
<th>Estimation mode (computation cost)</th>
<th>Estimation of covariance ( Q )?</th>
<th>Additional parameters?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariance inflation</td>
<td>Innovation statistics in the observation space</td>
<td>Method of moments</td>
<td>Online (low)</td>
<td>No (inflation ( \lambda ) instead)</td>
<td>Yes (inflation variance ( \sigma_1^2 ))</td>
</tr>
<tr>
<td>Lag-innovation Bayesian inference</td>
<td>Lag-innovation between consecutive times</td>
<td>Method of moments</td>
<td>Online (moderate)</td>
<td>Yes</td>
<td>Yes (temporal smoother ( \tau ))</td>
</tr>
<tr>
<td>Expectation-maximization</td>
<td>Innovation &amp; hyperparameter likelihoods</td>
<td>Likelihood methods</td>
<td>Online (high)</td>
<td>No (or joint parameter with ( R ))</td>
<td>Yes (prior distributions for ( \theta ))</td>
</tr>
<tr>
<td></td>
<td>Total likelihoods of the state-space model</td>
<td>Maximum likelihood</td>
<td>Offline (high)</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>
cently adapted with a Bayesian formulation in Liu et al. (2017) as well as Berry and Harlim (2017).

Perspectives

Beyond these possible improvements in techniques, we also discuss prospects for future work. A first perspective concerns the combination of methods, especially regarding the estimation modes, i.e., online or offline. In our opinion, a great advantage of the expectation-maximization algorithm is the absence of additional parameter and its robustness, due to the number of observations used to approximate the total state-space likelihood, potentially large when using an important batch period. Thus, it is a useful tool to get a first estimate and averaged form for $Q$ and $R$ matrices, i.e., to infer mean amplitudes and parametric shapes including bloc structures and spatial dependencies. Since the computational cost for this offline method is important, this calibration part has to be done once during the configuration stage of the assimilation system. Then, low-cost online methods, e.g., based on lag-innovations, can use the robust offline estimates as initial conditions and adapt to slow variations of $Q(k)$ and $R(k)$ using relevant parametric shapes of these matrices.

In a following work, we plan to evaluate offline and online methods to different cases, considering both constant and time varying $Q$ and $R$ matrices. First, we will compare the different methodologies to the linear and unidimensional model used in this review in order to evaluate their performance and convergence in the asymptotic case. We will use the root mean squared error and the coverage probability to measure the inference performance on mean and covariance. Then, we will test the estimation methods on the chaotic Lorenz-96 model and evaluate their performance and robustness varying the number of available observations. We will finally implement methodologies to a more realistic case for operational data assimilation, applying for instance to a mid-complexity general circulation model and real or simulated satellite data. In this case, the number of observations will be limited compared to the size of the state. To reduce the degrees of freedom, the use of parametric shapes for error covariances will be necessary to tackle the rank-deficient observations. Moreover, the use of deterministic ensemble filters with localizations will be necessary in such realistic data assimilation problems.

Finally, the Gaussian and additive formulation of the error terms generally stated in data assimilation as in Eqs (1-2) is extremely convenient. In practice, it allows the application of Kalman-based algorithm and greatly simplify the use of method of moments and maximum likelihood approaches as detailed in this review. But is it able to compensate for non-additive sources of errors? Indeed, in realistic data assimilation problems, errors are multiplicative or introduced in the model by misparametrization and/or parameter evolution. The presented methods have to be tested on those configurations to evaluate whether or not the additive and Gaussian formulation with covariances $Q$ and $R$ is robust enough and identify its limitations.

Acknowledgments. This work has been carried out as part of the Copernicus Marine Environment Monitoring Service (CMEMS) 3DA project. CMEMS is implemented by Mercator Ocean in the framework of a delegation agreement with the European Union. This work was also partially supported by FOCUS Establishing Supercomputing Center of Excellence. CERE is a member of Institut Pierre Simon Laplace (IPSL). A.C. has been funded by the project REDDA (#250711) of the Norwegian Research Council.

References


