

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 & Department of Meteorology, University of Reading, UK

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 \mathbf{Q} and observation \mathbf{R} error covariance 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 used for reconstructing the state, and several methods have been proposed since the 90s for their estimation. Some of them are based on the moments of various innovations, including those in the observation space and 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. It also discusses (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 dynamic and observation equations. This is statisti-

cally convenient and representative of a broad range 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 \mathbf{Q} and \mathbf{R} . Using the discrete time index k from 1 to K for the sake of simplicity, it is assumed that

$$\begin{cases} \mathbf{x}(k) = \mathcal{M}(k-1, \mathbf{x}(k-1)) + \boldsymbol{\eta}(k), & (1) \\ \mathbf{y}(k) = \mathcal{H}(k, \mathbf{x}(k)) + \boldsymbol{\epsilon}(k), & (2) \end{cases}$$

*Corresponding author address: Dept. Signal & Communications, IMT Atlantique, 655 Avenue du Technopole, 29200 Plouzané, France
E-mail: pierre.tandeo@imt-atlantique.fr

with \mathcal{M} the time dependent dynamic model, \mathcal{H} the time dependent transformation operator from the hidden state \mathbf{x} to the noisy observations \mathbf{y} , Gaussian white noise errors $\boldsymbol{\eta}(k) \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}(k))$ and $\boldsymbol{\epsilon}(k) \sim \mathcal{N}(\mathbf{0}, \mathbf{R}(k))$. We suppose that the initial state condition at $k = 0$ is a Gaussian vector with mean \mathbf{x}^b and variance \mathbf{B} and that $\boldsymbol{\eta}$ and $\boldsymbol{\epsilon}$ 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 \mathbf{x} conditionally to the observations \mathbf{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 reanalysis/smoothing highly depends on the assumed uncertainties associated to observations and to the model state, which have to be as realistic as possible. Using the formulation given in Eqs. (1-2), the uncertainties are represented through \mathbf{Q} and \mathbf{R} . In practice, the observation error covariance matrix \mathbf{R} in Eq. (2) can be determined empirically by estimating the instrument noise and the representativeness error between the state and the observation space, but a correct estimation of the latter is often challenging (Janjić et al. 2017). During the dynamic model evolution from $k - 1$ to k in Eq. (1), the model state is contaminated by two sources of uncertainty, the error in the state at $k - 1$ and the model uncertainty itself which is represented in state-space models via the additive model error term $\boldsymbol{\eta}$. Determining the model error covariance matrix \mathbf{Q} is difficult because it has to account for the model deficiencies 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 far fewer observations than those needed to estimate the entries of \mathbf{Q} (Daley 1992; Dee 1995).

When using either variational or ensemble-based DA methods, the quality of the reconstructed state vector largely depends on the relative amplitudes between the assumed observation and model errors. For instance, in Kalman filter-like methods, the ratio $\|\mathbf{Q}\|/\|\mathbf{R}\|$ 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 \mathbf{Q} , \mathbf{R} and $\|\mathbf{Q}\|/\|\mathbf{R}\|$ on the reconstruction of the state is not easy to evaluate. This is due to the complexity and size of the dynamic models, the effect of forcing terms and the huge variety of observations.

The importance of estimating error covariance matrices in Gaussian state-space models can be illustrated using a simple example with linear dynamics. Suppose that we aim at tracking a scalar state x governed by an autoregressive

sive AR(1) model in Eq. (1) defined by

$$x(k) = 0.95x(k-1) + \eta(k), \quad (3)$$

with $\eta \sim \mathcal{N}(0, Q^t)$ where the superscript t means “true” and $Q^t = 1$. Furthermore, observations y of the state are contaminated with another independent additive zero-mean and unit-variance Gaussian noise (i.e. $R^t = 1$) in Eq. (2) with $\mathcal{H}(x) = x$. 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 coefficient close to one and thus represents a process which evolves slowly in time. The linear dynamic model evolves stochastically and the measurement process also introduces a noise at each time step. Although the knowledge of these two sources of noise is crucial for the estimation problem, in practice identifying them is not an easy task. Given that the dynamic model is linear and the error terms are additive and Gaussian in this simple example, the Kalman smoother provides an exact algorithm 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 probability” or percentage of x falling in the 95% confidence intervals (defined as the mean ± 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 $Q^t = R^t = 1$ are respectively 0.71 and 95%. When using a low model error variance $Q = 0.1Q^t$ in Fig. 1(a), it gives a large weight to the forecasts given by the quasi-persistent autoregressive dynamic model. On the other hand, when using a low observation error variance $R = 0.1R^t$ in Fig. 1(b), excessive 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 as in 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 simultaneous underestimation and overestimation of model and observation errors. In both cases, the mean reconstructed state is equal to that obtained with

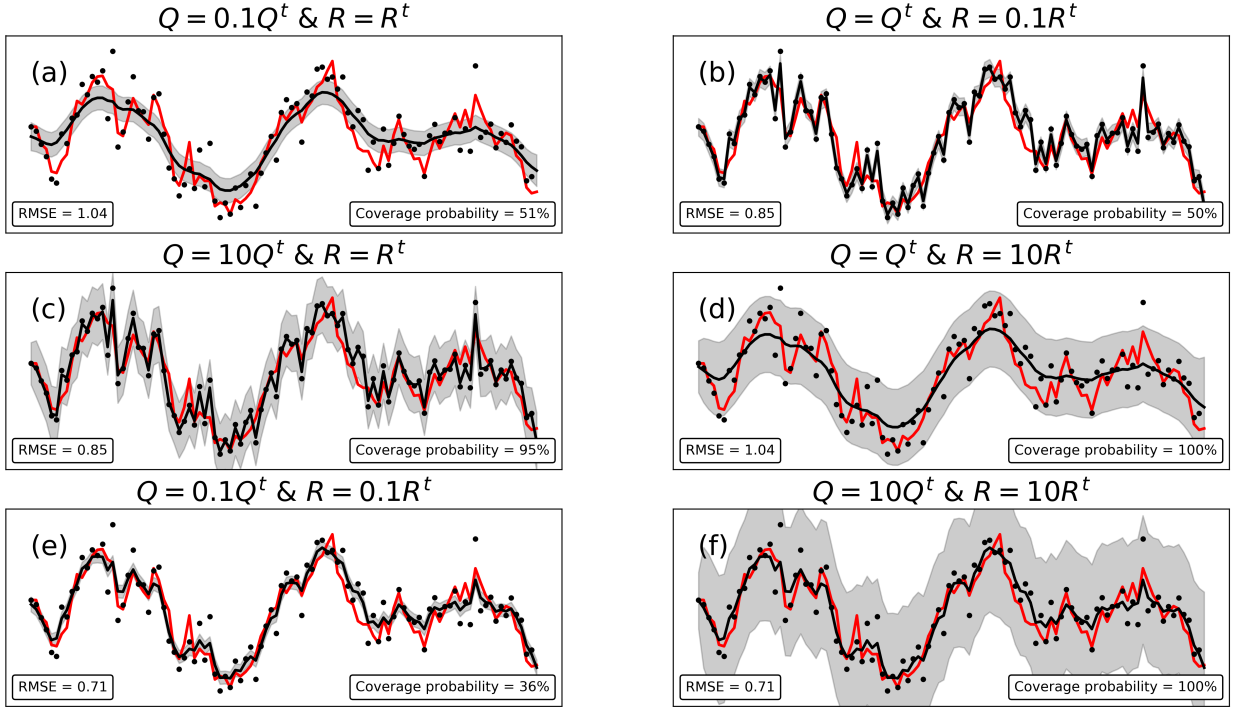


FIG. 1. Example of a univariate AR(1) process generated using Eq. (3) with $Q^t = 1$ (red line), noisy observations using Eq. (2) with $R^t = 1$ (black dots) and reconstructions using a Kalman smoother (black lines and gray 95% confidence interval) with different values of Q and R .

the true error variances, 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 once more the Kalman smoother where one of the variances Q or R is erroneous. This time, we compensate the error in the incorrectly prescribed variance by optimizing the other free variance using the maximum likelihood estimation method given in Shumway and Stoffer (1982), see Sect. 4b. Results presented in Fig. 2 show that best optimal RMSE (0.71) and coverage probability (95%) are reached close to the optimal variance noises $Q^t = R^t = 1$. 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, often only the RMSE is presented in research papers. The coverage probability is a measure of the DA capability to quantify 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 vary significantly as shown in Fig. 1(e) and Fig. 1(f).

Here, we use a simple synthetic example but for large dimensional and highly nonlinear dynamics, as in the DA in geosciences, such an underestimation or overestimation of uncertainty may have a strong impact and may make filters 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 90s, a significant number of works have dealt with error covariances in state-space models. The first to mention the importance of noise covariance matrices \mathbf{Q} and \mathbf{R} in DA were Hollingsworth and Lönnberg (1986), 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 covariance, denoted as \mathbf{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 \mathbf{Q} . Dee (1995) proposed a maximum likelihood estimator for parameterized versions of \mathbf{Q} and \mathbf{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). Dee (1995) showed that the maximum likelihood estimator is equivalent to the

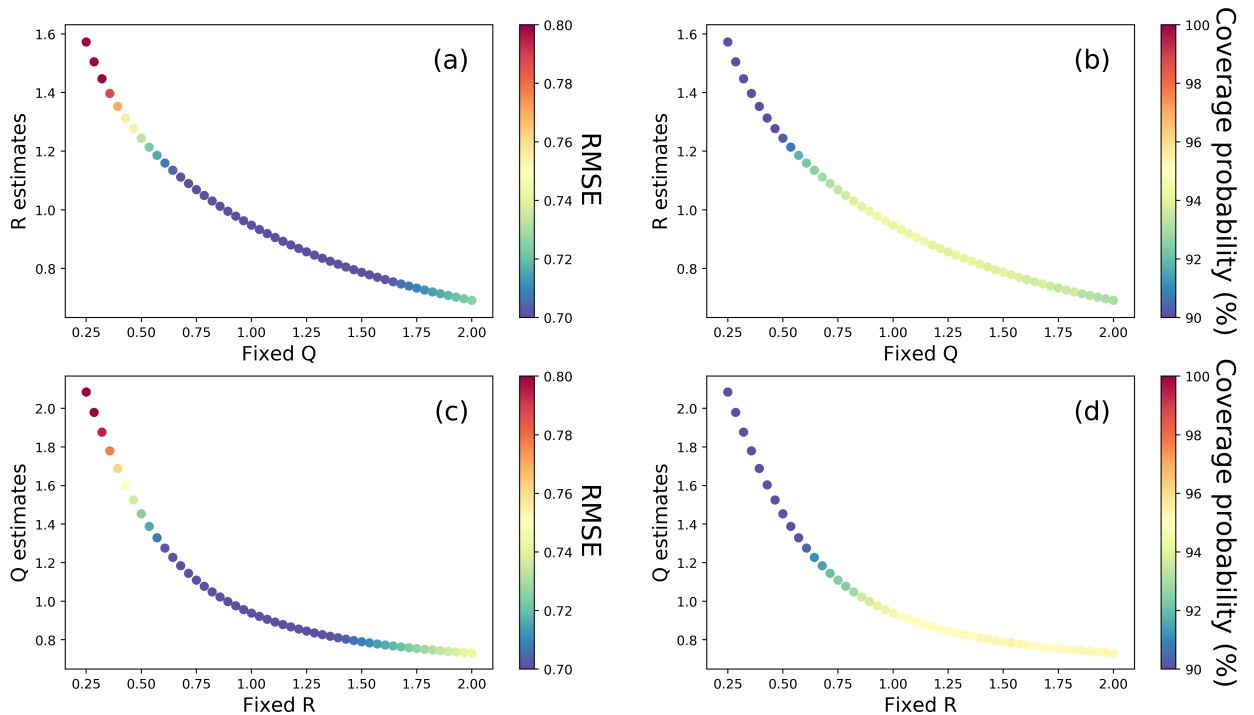


FIG. 2. Estimation results of R and Q for respectively fixed Q and R in the case of a Kalman smoother with the univariate AR(1) process given in Eq. (3). Here, we consider the maximum likelihood estimator given by Shumway and Stoffer (1982).

estimator given by the innovation covariances, hence making a strong connection between maximum likelihood and moment-based estimators. Desroziers and Ivanov (2001) used another estimation moment-based method built on an observation-minus-analysis diagnostic. Extending the early result by Dee (1995), it can be shown that the maximization of the innovation likelihood makes this diagnostic optimal (Chapnik et al. 2004), or in others terms, it implies that the χ^2 criterion is verified (Wu et al. 2013). 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 distinguish the impact of \mathbf{Q} and \mathbf{R} in the Kalman equations, and to estimate them independently is challenging in this case. Thus, they proposed various alternatives to tackle this issue.

A 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, we could classify the approaches whether they rely upon innovations or 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). This approach has been used extensively for the calibration of forecast covariance inflation with various implementations including additive, relaxation-to-prior and the multiplicative inflation case (Li et al. 2009a; Miyoshi 2011). Instead of working on different innovations at a given time, Berry and Sauer (2013) as well as Harlim et al. (2014) suggested using 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 such 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 estimation modes, for which the computational cost highly varies. In the online or adaptive approaches, one attempts to estimate time-dependent $\mathbf{Q}(k)$ and $\mathbf{R}(k)$ at the same time as the state vector, using filtering methods. When considering offline or batch approaches, averaged \mathbf{Q} and \mathbf{R} are

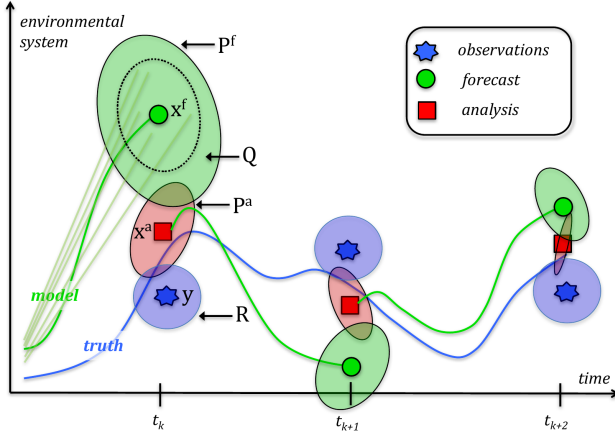


FIG. 3. Sketch of sequential data assimilation algorithms in the observation space where \mathbf{H} is omitted for simplicity reasons. The ellipses represent the error covariances computed from the Kalman-based equations, i.e. the forecast \mathbf{P}^f and analysis \mathbf{P}^a , as well as the model error \mathbf{Q} and observation error \mathbf{R} , unknown entries of the state-space model in Eqs. (1-2). This is a modified figure based on Fig. 1 from Carrasi et al. (2018).

estimated using all the observations within 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, based for instance on the likelihood. Finally, some methods requires tuning additional parameters and they have to be carefully chosen for practical implementations. We discuss this point later 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 Duník et al. (2017). The Mehra (1972) paper is a concise review which accounts for linear dynamic models using the classic Kalman filter while Kantas et al. (2015) and Duník et al. (2017) focus on the challenging covariance estimation in non-Gaussian Monte Carlo methods, and include various implementations of offline and online methods based on maximum likelihood and Bayesian approaches, but not methods based on innovation statistics. They evaluate the methods on one-dimensional and easy-to-simulate nonlinear models, using a large number of members or particles. The Kantas et al. (2015) and Duník et al. (2017) review papers address more theoretical aspects and asymptotic results. The current review is addressed to the DA community, so that we focus on methods which can be implemented for nonlinear operators and high dimensional systems, in particular using the Kalman formulation.

The review is organized as follows. Section 2 presents the filtering and smoothing DA algorithms used in this present work. The main families of methods used in the literature to jointly estimate error covariance matrices \mathbf{Q}

and \mathbf{R} are presented subsequently. First, innovation-based methods are presented in Sect. 3. Then, we describe in Sect. 4 the likelihood-based approaches. We also mention other alternatives in Sect. 5 with methods used in the past and not exactly matching the scope of this review, or diagnostic tools to check the accuracy of \mathbf{Q} and \mathbf{R} . Finally, in Sect. 6, we summarize this review and discuss possible perspectives for future works and remaining challenges to be faced in this domain.

2. Filtering and smoothing algorithms

For the overall discussion of the methods and for introduction of the notation, we present in this section a short description of the extended version of the Kalman equations for nonlinear dynamic systems and observation operators. Here we use time dependent linearizations \mathbf{M} and \mathbf{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, instead of the Ensemble Kalman Filter (EnKF) for instance, to avoid overburdening notations introduced by the ensemble members. However, the methods are also straightforward to apply in stochastic and square-root EnKFs (Houtekamer and Zhang 2016).

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. However, we focus in this review on Gaussian additive errors $\boldsymbol{\eta}$ and $\boldsymbol{\epsilon}$ 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 contention points (Atkins et al. 2013; Zhu et al. 2016). However, because 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 \mathbf{Q} with Gaussian Kalman-based methods may give a useful constraint or parameterization setup of the model error covariance matrix for PF (Zhu et al. 2017).

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

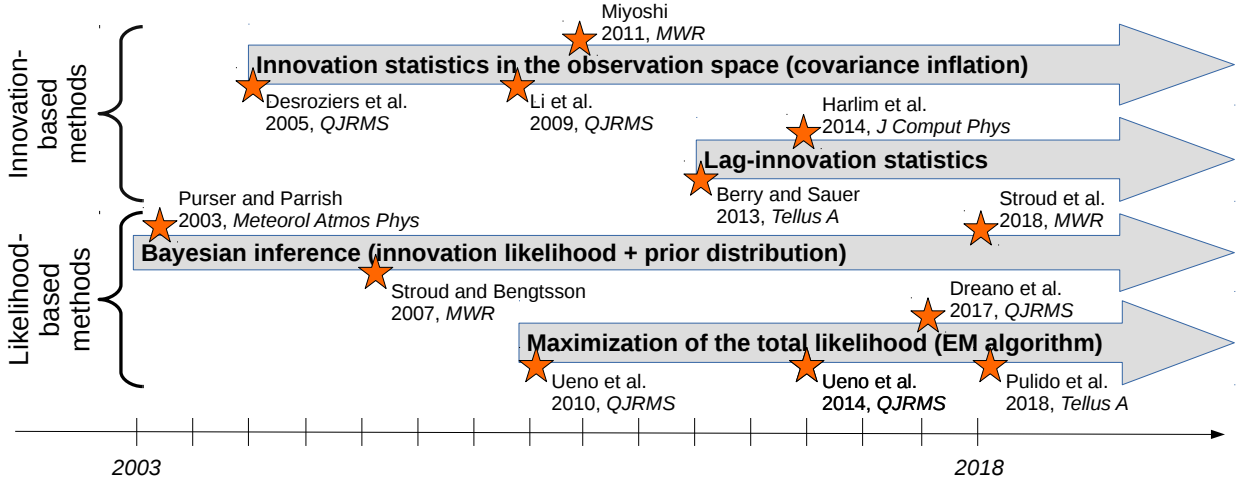


FIG. 4. Timeline of the main methodologies used in the data assimilation community for the joint estimation of \mathbf{Q} and \mathbf{R} in the last 15 years. Dee (1995) paper is not represented here but is certainly the seminal work of this research field in data assimilation.

Cosme et al. (2012). They are divided in three main steps:

Forecast step (forward in time):

$$\mathbf{x}^f(k) = \mathcal{M}(\mathbf{x}^a(k-1)) \quad (4)$$

$$\mathbf{P}^f(k) = \mathbf{M}(k)\mathbf{P}^a(k-1)\mathbf{M}(k)^\top + \mathbf{Q}(k) \quad (5)$$

Analysis step (forward in time):

$$\mathbf{d}(k) = \mathbf{y}(k) - \mathcal{H}(\mathbf{x}^f(k)) \quad (6)$$

$$\mathbf{K}^f(k) = \mathbf{P}^f(k)\mathbf{H}(k)^\top \left(\mathbf{H}(k)\mathbf{P}^f(k)\mathbf{H}(k)^\top + \mathbf{R}(k) \right)^{-1} \quad (7)$$

$$\mathbf{x}^a(k) = \mathbf{x}^f(k) + \mathbf{K}^f(k)\mathbf{d}(k) \quad (8)$$

$$\mathbf{P}^a(k) = (\mathbf{I} - \mathbf{K}^f(k)\mathbf{H}(k))\mathbf{P}^f(k) \quad (9)$$

Reanalysis step (backward in time):

$$\mathbf{K}^s(k) = \mathbf{P}^a(k)\mathbf{M}(k)^\top (\mathbf{P}^f(k+1))^{-1} \quad (10)$$

$$\mathbf{x}^s(k) = \mathbf{x}^a(k) + \mathbf{K}^s(k) (\mathbf{x}^s(k+1) - \mathbf{x}^f(k+1)) \quad (11)$$

$$\begin{aligned} \mathbf{P}^s(k) &= \mathbf{P}^a(k) \\ &\quad - \mathbf{K}^s(k) (\mathbf{P}^f(k+1) - \mathbf{P}^s(k+1)) \mathbf{K}^s(k)^\top \end{aligned} \quad (12)$$

$$\mathbf{P}^s(k, k+1) = \mathbf{P}^s(k+1)\mathbf{K}^s(k)^\top \quad (13)$$

Here, $\mathbf{P}^f(k)$, $\mathbf{P}^a(k)$, $\mathbf{P}^s(k)$ and $\mathbf{P}^s(k, k+1)$ denote respectively the covariance matrices of the forecast state $\mathbf{x}^f(k)$, the filtered state $\mathbf{x}^a(k)$, the smoother state $\mathbf{x}^s(k)$ and the pair $\{\mathbf{x}^s(k), \mathbf{x}^s(k+1)\}$. Finally, note that \mathbf{K}^f and \mathbf{K}^s are the filter and smoother Kalman gains and the innovation is denoted as \mathbf{d} .

3. Innovation-based methods

The importance of the innovation statistics has been emphasized in the DA community by Daley (1992), Dee (1995) and Desroziers and Ivanov (2001). The “classic innovation” \mathbf{d} , difference between the observations and the forecast states in the observation space, defined in Eq. (6), implicitly takes into account the \mathbf{Q} and \mathbf{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 \mathbf{Q} and \mathbf{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.

a. 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, $\mathbf{d}^{o-f}(k) = \mathbf{y}(k) - \mathcal{H}(\mathbf{x}^f(k))$ as in Eq. (6) and $\mathbf{d}^{o-a}(k) = \mathbf{y}(k) - \mathcal{H}(\mathbf{x}^a(k))$. We remark that another diagnostic using the difference between analysis $\mathbf{x}^a(k)$ and reanalysis $\mathbf{x}^s(k)$ has been proposed by Todling (2015) and

Bowler (2017) in the case of sequential and variational DA respectively to estimate the covariance \mathbf{Q} alone. In theory, in the linear and Gaussian case, the Desroziers innovation statistics should verify the equalities:

$$\begin{cases} \mathbb{E} [\mathbf{d}^{o-f}(k) \mathbf{d}^{o-f}(k)^\top] = \mathbf{H}(k) \mathbf{P}^f(k) \mathbf{H}(k)^\top + \mathbf{R}(k) \\ \mathbb{E} [\mathbf{d}^{o-a}(k) \mathbf{d}^{o-f}(k)^\top] = \mathbf{R}(k) \end{cases} \quad (15)$$

with \mathbb{E} the expectation operator. In this approach, we do not estimate \mathbf{Q} directly which is implicitly taken into account in \mathbf{P}^f . Instead, the approach attempts to compensate in \mathbf{P}^f for the lack of knowledge of \mathbf{Q} as well as 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), appendix A. Thus, covariance inflation can be required even in a perfect model scenario (i.e. with $\mathbf{Q} = \mathbf{0}$), 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 (Raanes et al. 2018).

We distinguish three inflation methods: multiplicative, additive and relaxation-to-prior. In the multiplicative case, the forecast error covariance matrix \mathbf{P}^f is usually multiplied by a scalar coefficient greater than 1 (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; 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 a 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 $\boldsymbol{\eta}$ in Eq. (1). Thus, it is not a direct estimation of \mathbf{Q} but an inflation applied to \mathbf{P}^f in order to compensate model uncertainties and sampling errors in EnKFs, as explained in Raanes et al. (2018), Sect. 4 and appendix C. Several DA systems work with an inflation method and used it for its simplicity, low-cost and efficiency.

The most straightforward online inflation estimation is a multiplicative factor λ of the badly scaled $\tilde{\mathbf{P}}^f(k)$ so that the corrected forecast covariance is given by $\mathbf{P}^f(k) = \lambda(k) \tilde{\mathbf{P}}^f(k)$. The estimate of the inflation factor is given by taking the trace of Eq. (14):

$$\lambda(k) = \mathbb{E} \left[\frac{\mathbf{d}^{o-f}(k)^\top \mathbf{d}^{o-f}(k) - \text{Tr}(\mathbf{R}(k))}{\text{Tr}(\mathbf{H}(k) \tilde{\mathbf{P}}^f(k) \mathbf{H}(k)^\top)} \right]. \quad (16)$$

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 σ_λ^2 . This parameter has to be carefully tuned to avoid filter divergence. Then, at each time step k , when sufficient observations are available, an estimate of $\mathbf{R}(k)$ is directly given by Eq. (15). Note that Li et al. (2009a) proposed to estimate each component of a diagonal \mathbf{R} matrix, and also suggested using an offline procedure to compute the average of these variance terms.

b. Lag-innovation between consecutive times

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

$$\begin{cases} \mathbb{E} [\mathbf{d}(k) \mathbf{d}(k)^\top] = \mathbf{H}(k) \mathbf{P}^f(k) \mathbf{H}(k)^\top + \mathbf{R}(k) = \boldsymbol{\Sigma}(k) \\ \mathbb{E} [\mathbf{d}(k) \mathbf{d}(k-1)^\top] = \mathbf{H}(k) \mathbf{M}(k) \mathbf{P}^f(k-1) \mathbf{H}(k-1)^\top \\ - \mathbf{H}(k) \mathbf{M}(k) \mathbf{K}^f(k-1) \boldsymbol{\Sigma}(k-1). \end{cases} \quad (18)$$

Lag-innovations were introduced by Mehra (1970) in order to simultaneously recover the error covariance matrices for a Gaussian and linear state-space model. Mehra established analytic exact relations between \mathbf{Q} and \mathbf{R} , and the probabilistic expectations of $\mathbf{d}(k) \mathbf{d}(k-l)^\top$ for linear systems in steady state. Then, Bélanger (1974) extended these results to the case of time-varying linear stochastic processes, taking $\mathbf{d}(k) \mathbf{d}(k-l)^\top$ as “observations” of \mathbf{Q} and \mathbf{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 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 Mehra’s method and Harlim

et al. (2014) followed the original Bélanger 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 describe the algorithm of Berry and Sauer (2013) using lag-one innovations. It is based on the online (or adaptive) estimation of $\mathbf{Q}(k)$ and $\mathbf{R}(k)$, which satisfies the following relations in the linear and Gaussian case:

$$\tilde{\mathbf{P}}(k) = \mathbf{M}(k)^{-1} \mathbf{H}(k)^{-1} \mathbf{d}(k) \mathbf{d}(k-1)^\top \mathbf{H}(k)^{-\top} + \mathbf{K}^f(k-1) \mathbf{d}(k-1) \mathbf{d}(k-1)^\top \mathbf{H}(k)^{-\top} \quad (19)$$

$$\tilde{\mathbf{Q}}(k) = \tilde{\mathbf{P}}(k) - \mathbf{M}(k-1) \mathbf{P}^d(k-1) \mathbf{M}(k-1)^\top \quad (20)$$

$$\tilde{\mathbf{R}}(k) = \mathbf{d}(k-1) \mathbf{d}(k-1)^\top - \mathbf{H}(k) \mathbf{P}^f(k-1) \mathbf{H}(k)^\top. \quad (21)$$

In this online procedure, joint estimations of $\tilde{\mathbf{Q}}(k)$ and $\tilde{\mathbf{R}}(k)$ 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,

$$\mathbf{Q}(k+1) = \mathbf{Q}(k) + (\tilde{\mathbf{Q}}(k) - \mathbf{Q}(k))/\tau, \quad (22)$$

$$\mathbf{R}(k+1) = \mathbf{R}(k) + (\tilde{\mathbf{R}}(k) - \mathbf{R}(k))/\tau \quad (23)$$

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

It is worth pointing out that in the case of sparse observations, the estimate of $\tilde{\mathbf{P}}$ in Eq. (19) might be underdetermined, even if the system is observable (Cohn and Dee 1988). This is attributed to the use of only one lag-innovation. Theoretically, all components of \mathbf{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 model defined in Lorenz (1996).

4. Likelihood-based methods

Approaches based on the likelihood function were put forward in the DA community by Dee (1995), Blanchet et al. (1997) as well as by Mitchell and Houtekamer (2000) where it was proposed to maximize the likelihood of the innovation, i.e. $p(\mathbf{y}(k)|\mathbf{y}(k-1))$, defined by the mean vector $\mathbf{d}(k)$ computed in Eq. (6) and covariance matrix $\Sigma(k)$ introduced in Eq. (17) and also used in the computation of the Kalman filter gain in Eq. (7). Unfortunately,

they reach the same conclusions as for the innovation-based methods, i.e. the joint estimation of \mathbf{Q} and \mathbf{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 \mathbf{Q} and \mathbf{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 several time steps 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 \mathbf{Q} and \mathbf{R} covariance matrices have *a priori* distributions which are controlled by certain hyperparameters. In practice, it is difficult to have a prior distribution for each element of \mathbf{Q} and \mathbf{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 $\boldsymbol{\theta}$. Then, we jointly and adaptively estimate the state \mathbf{x} and parameters $\boldsymbol{\theta}$ using Bayes' theorem:

$$p(\mathbf{x}(k), \boldsymbol{\theta}(k) | \mathbf{y}(1:k)) = p(\mathbf{x}(k) | \mathbf{y}(1:k), \boldsymbol{\theta}(k)) p(\boldsymbol{\theta}(k) | \mathbf{y}(1:k)). \quad (24)$$

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

$$p(\boldsymbol{\theta}(k) | \mathbf{y}(1:k)) \propto p(\mathbf{y}(k) | \mathbf{y}(1:k-1), \boldsymbol{\theta}(k)) p(\boldsymbol{\theta}(k) | \mathbf{y}(1:k-1)). \quad (25)$$

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 \mathbf{Q} , assuming that \mathbf{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 \mathbf{Q} and \mathbf{R} . In that case, the scalar parameter affects simultaneously the \mathbf{Q} and \mathbf{R} matrices. Based on the experiments about the importance of $\|\mathbf{Q}\|/\|\mathbf{R}\|$ ratio presented in Fig. 1, we can guess that this approach may not be optimal. Then, other works have applied similar Bayesian approaches for the estimation of parameters governing the shape of \mathbf{R} only: Frei and Künsch (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} for simple systems and only the estimation of \mathbf{R} for the Lorenz-96 model.

The Bayesian inference approach is an online estimation procedure with joint estimation of system's state and hyperparameters $\boldsymbol{\theta}(k)$, embedding 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 require a large number of Monte-Carlo simulations to estimate correctly $p(\boldsymbol{\theta}(k)|\mathbf{y}(1:k))$ defined in Eq. (25) or iterative procedures as in Ueno and Nakamura (2016). Alternatively, Scheffler et al. (2018) assume a Gaussian distribution for $p(\boldsymbol{\theta}(k)|\mathbf{y}(1:k-1))$ 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, but recent works have so far assumed that $\boldsymbol{\theta}(k)$ does not depend on 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 in Stroud et al. (2018).

b. Maximization of the total likelihoods

The innovation likelihood at time k is defined by $p(\mathbf{y}(k)|\mathbf{y}(1:k-1), \boldsymbol{\theta}(k))$ in Eq. (25). Maximizing this likelihood at each time step has been proposed by various authors: Dee (1995), Blanchet 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 $\boldsymbol{\Sigma}(k) = \mathbf{H}(k)\mathbf{P}^f(k)\mathbf{H}(k)^\top + \mathbf{R}(k)$ combines the information of \mathbf{R} and \mathbf{Q} , the latter contained in \mathbf{P}^f . When using only time k , it is difficult to distinguish the model and observation error covariances and in practice, the aforementioned works 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 parameters by two or three orders of magnitude”. A reasonable alternative is to use a batch of observations distributed in time and to assume $\boldsymbol{\theta}$ to be constant in time. The resulting total likelihood expressed sequentially through conditioning is

given by

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

This likelihood is said to be “incomplete” because it only depends on the observations of the state-space system, not on the hidden state. But since it is an integration of innovation likelihoods over a long period of time, this provides 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 Carrasi 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 tested 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

$$p(\mathbf{y}(1:K), \mathbf{x}(0:K)|\boldsymbol{\theta}) = p(\mathbf{x}(0)) \prod_{k=1}^K p(\mathbf{x}(k)|\mathbf{x}(k-1), \boldsymbol{\theta}) \prod_{k=1}^K p(\mathbf{y}(k)|\mathbf{x}(k), \boldsymbol{\theta}) \quad (27)$$

where the three terms on the rhs are related to the initial state, to the state equation in Eq. (1) and to the observation equation in Eq. (2). In practice, the marginalization of the full hidden state from $k = 0$ to K is not possi-

ble, 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 maximize the likelihood criterion given in Eq. (27). They used the Expectation-Maximization algorithm (hereinafter noted EM) introduced by Dempster et al. (1977) to ensure the convergence to the maximum likelihood estimator, and applied it to estimate \mathbf{Q} and \mathbf{R} in the case of linear dynamics. In DA, the EM algorithm has been implemented for estimating only \mathbf{R} in Ueno and Nakamura (2014) for the Zebiak and Cane (ZC) model and satellite altimetry, for the background covariance matrix \mathbf{B} and \mathbf{R} in Liu et al. (2017) in case of accidental pollutant source retrieval, for both \mathbf{Q} and \mathbf{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 it to a gradient ascent optimization of the incomplete total likelihood to estimate \mathbf{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 \mathbf{Q} and \mathbf{R} and show the improvement compared to Kalman-based filters.

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

$$\begin{aligned} \mathbf{Q} = & \frac{1}{K} \sum_{k=1}^K \{ (\mathbf{x}^s(k) - \mathcal{M}\mathbf{x}^s(k-1)) (\mathbf{x}^s(k) - \mathcal{M}\mathbf{x}^s(k-1))^\top \\ & + \mathbf{M}(k)\mathbf{P}^s(k-1)\mathbf{M}(k)^\top + \mathbf{P}^s(k) \\ & - \mathbf{P}^s(k-1, k)\mathbf{M}(k)^\top - \mathbf{M}(k)\mathbf{P}^s(k-1, k)^\top \} \end{aligned} \quad (28)$$

$$\begin{aligned} \mathbf{R} = & \frac{1}{K} \sum_{k=1}^K \{ (\mathbf{y}(k) - \mathbf{H}(k)\mathbf{x}^s(k)) (\mathbf{y}(k) - \mathbf{H}(k)\mathbf{x}^s(k))^\top \\ & + \mathbf{H}(k)\mathbf{P}^s(k)\mathbf{H}(k)^\top \}. \end{aligned} \quad (29)$$

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 high because it requires the use of Kalman-based filter and smoother at each iteration. Nevertheless, this EM method does not require any additional parameter and is robust to initial conditions, i.e. the values given for \mathbf{Q} and \mathbf{R} matrices in the first EM iteration. Additionally, note that the maximization of the total likelihood allows the estimation of the initial state vector \mathbf{x}^b and covariance matrix \mathbf{B} as discussed in Tandeo et al. (2015) and Dreano et al. (2017). In that case, we should write $p(\mathbf{x}(0)|\theta)$ instead of $p(\mathbf{x}(0))$ 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 problems 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 covariances 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 (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 \mathbf{Q} and \mathbf{R} . Furthermore, another critical aspect of this approach is that one needs to define an evolution model for the augmented state. 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. 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 increment refers to statistical methods that study the relationship between two consecutive times of a dynamic system. The use of regression methods has been firstly proposed by Lorenz (1977) and then by Leith (1978) to learn error statistics of dynamic models in meteorology (e.g. bias and covariance \mathbf{Q} of the error $\boldsymbol{\eta}$). 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 Vanitsem (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. χ^2 test, reliability budget, cross-validation, whiteness of lag-innovations

These methods are not direct estimations 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 χ^2 distribution with a given number of degrees of freedom. As pointed out respectively by Michalak et al. (2005) and Rayner et al. (1999), lots of combination of \mathbf{Q} and \mathbf{R} errors can lead to the acceptance of this test and this method cannot guide the relative allocation of error between the two. The second one is based on the "reliability budget" in ensemble DA, decomposed in 3 terms including innovation bias, \mathbf{P}^f and \mathbf{R} . Each term is averaged in time, and significance is tested using Student's t-tests (Rodwell et al. 2016). 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 Wahba et al. (1995) or Wu et al. (2013) for more details. The last method was raised 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 \mathbf{Q} and \mathbf{R} matrices.

6. Summary, conclusions and perspectives

In this review paper, we presented different methods to estimate 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 \mathbf{Q} and \mathbf{R} . The individual and joint impacts of badly calibrated covariances

were firstly explained using a linear toy-model. The experiments clearly showed that to achieve 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 \mathbf{Q} and \mathbf{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 forecast state, and the use of empirical and theoretical moments, also known as the method 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 \mathbf{Q} is not necessarily explicit and where an estimated inflation factor artificially increases the forecast error covariance \mathbf{P}^f . This online method is low-cost, adaptive, used in many operational data assimilation systems with different implementations like the multiplicative case (Raanes et al. 2018). 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, while 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 τ , 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. So far, authors considered constant \mathbf{Q} and \mathbf{R} but in practice, when τ 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 \mathbf{Q} and \mathbf{R} using only current observations is limited. Thus, some 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

TABLE 1. Comparison of several methods to estimate error covariances matrices \mathbf{Q} and \mathbf{R} in data assimilation.

Method	Criteria	Estimation method	Estimation mode (computation cost)	Estimation of covariance \mathbf{Q}	Tunable parameters
Covariance inflation	Innovation statistics in the observation space	Method of moments	Online (low)	No (inflation λ instead)	Yes (inflation variance σ_λ^2)
Lag-innovation	Lag-innovation between consecutive times	Method of moments	Online (moderate)	Yes	Yes (temporal smoother τ)
Bayesian inference	Innovation & hyperparameter likelihoods	Likelihood methods	Online (high)	No (or joint parameter with \mathbf{R})	Yes (prior distributions for θ)
Expectation-maximization	Total likelihoods of the state-space model	Maximum likelihood	Offline (high)	Yes	No

that have to be carefully chosen. Moreover, large Monte-Carlo simulations are needed to estimate the hyperparameters of these distributions. The joint estimation of \mathbf{Q} and \mathbf{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 in 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 by Shumway and Stoffer (1982). Given its stability and robustness, we choose to detail the EM procedure that iteratively aims at maximizing the complete total likelihood, even in nonlinear 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 \mathbf{Q} and \mathbf{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 \mathbf{R} in a realistic coupled atmosphere-ocean model and Liu et al. (2017) applied the 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 \mathbf{Q} and \mathbf{R} estimates in time. Instead of using fixed values for these

parameters, for instance τ in lag-innovations or σ_λ^2 in inflation methods, we suggest using 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 (τ high, σ_λ^2 low) and iterate the procedures on a batch of observations as in the EM algorithm. This was suggested and tested in Desroziers et al. (2005) with encouraging results. To the best of our knowledge, it has not yet been tested with lag-innovation methods.

The second challenge concerns the Bayesian approach where joint estimation of \mathbf{Q} and \mathbf{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 \mathbf{Q} and \mathbf{R} might physically constrain 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. Thus, the \mathbf{Q} and \mathbf{R} could be smoothed in time. Another way is to apply online EM algorithms (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 (Pulido et al. 2018).

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 conditions \mathbf{x}^b and \mathbf{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 η and ϵ . 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 recently 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 EM algorithm is the absence of additional parameters and its robustness, due to the number of observations used to approximate the total state-space likelihood, potentially large when using a substantial batch period. Thus, it is a useful tool to get a first estimate and averaged form for \mathbf{Q} and \mathbf{R} matrices, i.e. to infer mean amplitudes and parametric shapes including block structures and spatial dependencies. Since the computational cost for this offline method is significant, 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 $\mathbf{Q}(k)$ and $\mathbf{R}(k)$ 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 \mathbf{Q} and \mathbf{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 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 by varying the number of available observations. We will finally implement methodologies to a more realistic case for operational data assimilation, applied 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 necessarily to tackle the rank-deficient observations. Moreover, the use of deterministic ensemble filters with localizations will be necessarily in such realistic data assimilation problems (Houtekamer and Zhang 2016).

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 simplifies 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 into 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 \mathbf{Q} and \mathbf{R} is robust enough and also to 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. CEREIA is a member of Institut Pierre Simon Laplace (IPSL). A.C. has been funded by the project REDDA (#250711) of the Norwegian Research Council. We thank John C. Wells for his English grammar corrections.

References

- Anderson, J. L., and S. L. Anderson, 1999: A Monte Carlo Implementation of the Nonlinear Filtering Problem to Produce Ensemble Assimilations and Forecasts. *Monthly Weather Review*, **127**, 2741–2758.
- Atkins, E., M. Morzfeld, and A. J. Chorin, 2013: Implicit Particle Methods and Their Connection with Variational Data Assimilation. *Monthly Weather Review*, **141** (6), 1786–1803.
- Bélanger, P. R., 1974: Estimation of noise covariance matrices for a linear time-varying stochastic process. *Automatica*, **10** (3), 267–275.
- Berry, T., and J. Harlim, 2017: Correcting Biased Observation Model Error in Data Assimilation. *Monthly Weather Review*, **145** (7), 2833–2853.
- Berry, T., and T. Sauer, 2013: Adaptive ensemble Kalman filtering of non-linear systems. *Tellus, Series A: Dynamic Meteorology and Oceanography*, **65** (20331), 1–16.
- Berry, T., and T. Sauer, 2018: Correlation between system and observation errors in data assimilation. *Monthly Weather Review*, **146** (9), 2913–2931.
- Blanchet, I., C. Frankignoul, and M. A. Cane, 1997: A comparison of adaptive kalman filters for a tropical pacific ocean model. *Monthly Weather Review*, **125** (1), 40–58.
- Bocquet, M., 2011: Ensemble Kalman filtering without the intrinsic need for inflation. *Nonlinear Processes in Geophysics*, **18** (5), 735–750.
- Bocquet, M., and P. Sakov, 2012: Combining inflation-free and iterative ensemble Kalman filters for strongly nonlinear systems. *Nonlinear Processes in Geophysics*, **19**, 383–399.
- Bowler, N. E., 2017: On the diagnosis of model error statistics using weak-constraint data assimilation. *Quarterly Journal of the Royal Meteorological Society*, **143** (705), 1916–1928.
- Brankart, J.-M., E. Cosme, C.-E. Testut, P. Brasseur, and J. Verron, 2010: Efficient Adaptive Error Parameterizations for Square Root

- or Ensemble Kalman Filters: Application to the Control of Ocean Mesoscale Signals. *Monthly Weather Review*, **138** (3), 932–950.
- Cappé, O., 2011: Online Expectation-Maximisation. *Mixtures: Estimation and Applications*, Wiley Series in Probability and Statistics, 1–53.
- Carrassi, A., M. Bocquet, L. Bertino, and G. Evensen, 2018: Data Assimilation in the Geosciences: An overview on methods, issues and perspectives. *WIREs Clim Change*, **9** (5), e535, doi:https://doi.org/10.1002/wcc.535.
- Carrassi, A., M. Bocquet, A. Hannart, and M. Ghil, 2017: Estimating model evidence using data assimilation. *Quarterly Journal of the Royal Meteorological Society*, **143** (703), 866–880.
- Carrassi, A., and S. Vannitsem, 2010: Accounting for Model Error in Variational Data Assimilation: A Deterministic Formulation. *Monthly Weather Review*, **138** (9), 3369–3386.
- Carrassi, A., and S. Vannitsem, 2011a: State and parameter estimation with the extended Kalman filter: An alternative formulation of the model error dynamics. *Quarterly Journal of the Royal Meteorological Society*, **137** (1), 435–451.
- Carrassi, A., and S. Vannitsem, 2011b: Treatment of the Error Due To Unresolved Scales in Sequential Data Assimilation. *International Journal of Bifurcation and Chaos*, **21** (12), 3619–3626.
- Chapnik, B., G. Desroziers, F. Rabier, and O. Talagrand, 2004: Properties and first application of an error-statistics tuning method in variational assimilation. *Quarterly Journal of the Royal Meteorological Society*, **130** (601), 2253–2275.
- Chau, T. T. T., P. Ailliot, V. Monbet, and P. Tandeo, 2018: An efficient particle-based method for maximum likelihood estimation in nonlinear state-space models. URL <http://arxiv.org/abs/1804.07483>.
- Cohn, S. E., and D. P. Dee, 1988: Observability of Discretized Partial Differential Equations. *SIAM Journal on Numerical Analysis*, **25** (3), 586–617.
- Corazza, M., E. Kalnay, D. J. Patil, R. Morss, M. Cai, I. Szunyogh, B. R. Hunt, and J. A. Yorke, 2003: Use of the breeding technique to estimate the structure of the analysis errors of the day. *Nonlinear Processes in Geophysics*, **10** (3), 233–243.
- Cosme, E., J. Verron, P. Brasseur, J. Blum, and D. Auroux, 2012: Smoothing Problems in a Bayesian Framework and Their Linear Gaussian Solutions. *Monthly Weather Review*, **140** (2), 683–695.
- Daley, R., 1991: Atmospheric data analysis. URL <http://www.cambridge.org/catalogue/catalogue.asp?isbn=9780521458252>, 457 pp.
- Daley, R., 1992: Estimating Model-Error Covariances for Application to Atmospheric Data Assimilation. *Monthly Weather Review*, **120** (8), 1735–1746.
- Dee, D. P., 1995: On-line Estimation of Error Covariance Parameters for Atmospheric Data Assimilation. *Monthly Weather Review*, **123** (4), 1128–1145.
- Dee, D. P., S. E. Cohn, A. Dalcher, and M. Ghil, 1985: An Efficient Algorithm for Estimating Noise Covariances in Distributed Systems. *IEEE Transactions on Automatic Control*, **30** (11), 1057–1065.
- Dee, D. P., G. Gaspari, C. Redder, L. Rukhovets, and A. M. da Silva, 1999a: Maximum-Likelihood Estimation of Forecast and Observation Error Covariance Parameters. Part I: Methodology. *Monthly Weather Review*, **127** (1992), 1822–1834.
- Dee, D. P., G. Gaspari, C. Redder, L. Rukhovets, and A. M. da Silva, 1999b: Maximum-Likelihood Estimation of Forecast and Observation Error Covariance Parameters. Part II: Applications. *Monthly Weather Review*, **127** (1992), 1835–1849.
- Delsole, T., and X. Yang, 2010: State and parameter estimation in stochastic dynamical models. *Physica D: Nonlinear Phenomena*, **239** (18), 1781–1788.
- Dempster, A. P., N. M. Laird, and D. B. Rubin, 1977: Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, **39** (1), 1–38.
- Desroziers, G., L. Berre, B. Chapnik, and P. Poli, 2005: Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society*, **131** (613), 3385–3396.
- Desroziers, G., and S. Ivanov, 2001: Diagnosis and adaptive tuning of observation-error parameters in a variational assimilation. *Quarterly Journal of the Royal Meteorological Society*, **127** (574), 1433–1452.
- Dreano, D., P. Tandeo, M. Pulido, T. Chonavel, B. Alt-El-Fquih, and I. Hoteit, 2017: Estimating model error covariances in nonlinear state-space models using Kalman smoothing and the expectation-maximisation algorithm. *Quarterly Journal of the Royal Meteorological Society*, **143** (705), 1877–1885.
- Duník, J., O. Straka, O. Kost, and J. Havlík, 2017: Noise covariance matrices in state-space models: A survey and comparison of estimation methods-Part I. *International Journal of Adaptive Control and Signal Processing*, **31** (11), 1505–1543.
- Frei, M., and H. R. Künsch, 2012: Sequential State and Observation Noise Covariance Estimation Using Combined Ensemble Kalman and Particle Filters. *Monthly Weather Review*, **140** (5), 1476–1495.
- Fu, L.-L., I. Fukumori, and R. N. Miller, 1993: Fitting Dynamic Models to the Geosat Sea Level Observations in the Tropical Pacific Ocean. Part II: A Linear, Wind-driven Model. *Journal of Physical Oceanography*, **23** (10), 2162–2181.
- Ghil, M., and P. Malanotte-Rizzoli, 1991: Data assimilation in meteorology and oceanography.pdf. *Advances in Geophysics*, **33**, 141–266.
- Gordon, N., D. Salmond, and A. Smith, 1993: Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEEE Proceedings F: Radar and Signal Processing*, **140** (2), 107.
- Hannart, A., A. Carrassi, M. Bocquet, M. Ghil, P. Naveau, M. Pulido, J. Ruiz, and P. Tandeo, 2016: DADA: data assimilation for the detection and attribution of weather and climate-related events. *Climatic Change*, **136** (2).
- Harlim, J., A. Mahdi, and A. J. Majda, 2014: An ensemble Kalman filter for statistical estimation of physics constrained nonlinear regression models. *Journal of Computational Physics*, **257**, 782–812.
- Hollingsworth, A., and P. Lönnberg, 1986: The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. *Tellus A*, **38** (2), 111–136.

- Houtekamer, P. L., H. L. Mitchell, and X. Deng, 2009: Model Error Representation in an Operational Ensemble Kalman Filter. *Monthly Weather Review*, **137** (7), 2126–2143.
- Houtekamer, P. L., and F. Zhang, 2016: Review of the Ensemble Kalman Filter for Atmospheric Data Assimilation. *Monthly Weather Review*, **144** (12), 4489–4532.
- Janjić, T., and Coauthors, 2017: On the representation error in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, doi:10.1002/qj.3130, URL <http://doi.wiley.com/10.1002/qj.3130>.
- Jazwinski, A. H., 1970: *Stochastic processes and filtering theory*. Academic Press.
- Kantas, N., A. Doucet, S. S. Singh, J. Maciejowski, and N. Chopin, 2015: On Particle Methods for Parameter Estimation in State-Space Models. *Statistical Science*, **30** (3), 328–351.
- Kotsuki, S., Y. Ota, and T. Miyoshi, 2017: Adaptive covariance relaxation methods for ensemble data assimilation: Experiments in the real atmosphere. *Quarterly Journal of the Royal Meteorological Society*, **143** (705), 2001–2015.
- Leith, C. E., 1978: Objective Methods for Weather Prediction. *Annual Review of Fluid Mechanics*, **10** (1), 107–128.
- Li, H., E. Kalnay, and T. Miyoshi, 2009a: Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. *Quarterly Journal of the Royal Meteorological Society*, **135** (2), 523–533.
- Li, H., E. Kalnay, T. Miyoshi, and C. M. Danforth, 2009b: Accounting for Model Errors in Ensemble Data Assimilation. *Monthly Weather Review*, **137** (10), 3407–3419.
- Liang, X., X. Zheng, S. Zhang, G. Wu, Y. Dai, and Y. Li, 2012: Maximum likelihood estimation of inflation factors on error covariance matrices for ensemble Kalman filter assimilation. *Quarterly Journal of the Royal Meteorological Society*, **138** (662), 263–273.
- Liu, Y., J.-M. Haussaire, M. Bocquet, Y. Roustan, O. Saunier, and A. Mathieu, 2017: Uncertainty quantification of pollutant source retrieval: comparison of Bayesian methods with application to the Chernobyl and Fukushima Daiichi accidental releases of radionuclides. *Quarterly Journal of the Royal Meteorological Society*, **143** (708), 2886–2901.
- Lorenz, E. N., 1977: An Experiment in Nonlinear Statistical Weather Forecasting. *Monthly Weather Review*, **105**, 590–602.
- Lorenz, E. N., 1996: Predictability: A problem partly solved. *Proc. Seminar on predictability*, Vol. 1.
- Lorenz, E. N., and K. a. Emanuel, 1998: Optimal Sites for Supplementary Weather Observations: Simulation with a Small Model. *Journal of the Atmospheric Sciences*, **55** (3), 399–414.
- Mehra, R. K., 1970: On the Identification of Variances and Adaptive Kalman Filtering. *IEEE Transactions on Automatic Control*, **AC-15** (2), 175–184.
- Mehra, R. K., 1972: Approaches to Adaptive Filtering. *IEEE Transactions on Automatic Control*, **17** (5), 693–698.
- Menemenlis, D., and M. Chechelnitsky, 2000: Error Estimates for an Ocean General Circulation Model from Altimeter and Acoustic Tomography Data. *Monthly Weather Review*, **128** (3), 763–778.
- Michalak, A. M., A. Hirsch, L. Bruhwiler, K. R. Gurney, W. Peters, and P. P. Tans, 2005: Maximum likelihood estimation of covariance parameters for Bayesian atmospheric trace gas surface flux inversions. *Journal of Geophysical Research*, **110** (D24), D24 107.
- Mitchell, H. L., and P. L. Houtekamer, 2000: An adaptive ensemble Kalman filter. *Monthly Weather Review*, **128** (2), 416–433.
- Mitchell, L., and A. Carrassi, 2015: Accounting for model error due to unresolved scales within ensemble Kalman filtering. *Quarterly Journal of the Royal Meteorological Society*, **141** (689), 1417–1428.
- Miyoshi, T., 2011: The Gaussian Approach to Adaptive Covariance Inflation and Its Implementation with the Local Ensemble Transform Kalman Filter. *Monthly Weather Review*, **139** (5), 1519–1535.
- Pulido, M., P. Tandeo, M. Bocquet, A. Carrassi, and M. Lucini, 2018: Stochastic parameterization identification using ensemble Kalman filtering combined with maximum likelihood methods. *Tellus A: Dynamic Meteorology and Oceanography*, **70** (1), 1442 099.
- Purser, R. J., and D. F. Parrish, 2003: A Bayesian technique for estimating continuously varying statistical parameters of a variational assimilation. *Meteorology and Atmospheric Physics*, **82** (1-4), 209–226.
- Raanes, P. N., M. Bocquet, and A. Carrassi, 2018: Adaptive covariance inflation in the ensemble Kalman filter by Gaussian scale mixtures. *Quarterly Journal of the Royal Meteorological Society*, Accepted for publication, URL <http://arxiv.org/pdf/1801.08474.pdf>.
- Rayner, P. J., I. G. Enting, R. J. Francey, and R. Langenfelds, 1999: Reconstructing the recent carbon cycle from atmospheric CO₂, $\delta^{13}\text{C}$ and O₂/N₂ observations. *Tellus, Series B: Chemical and Physical Meteorology*, **51** (2), 213–232.
- Rodwell, M. J., S. T. Lang, N. B. Ingleby, N. Bormann, E. Hólm, F. Rabier, D. S. Richardson, and M. Yamaguchi, 2016: Reliability in ensemble data assimilation. *Quarterly Journal of the Royal Meteorological Society*, **142** (694), 443–454.
- Ruiz, J. J., M. Pulido, and T. Miyoshi, 2013: Estimating Model Parameters with Ensemble-Based Data Assimilation: A Review. *Journal of the Meteorological Society of Japan. Ser. II*, **91** (2), 79–99.
- Scheffler, G., J. Ruiz, and M. Pulido, 2018: Inference of stochastic parameterizations for model error treatment using nested ensemble Kalman filters. URL <https://arxiv.org/pdf/1807.10858.pdf>.
- Schmidt, S. F., 1966: Applications of state space methods to navigation problems. *Advances in Control Systems*, **3**, 293–340.
- Shumway, R. H., and D. S. Stoffer, 1982: An Approach to Time Series Smoothing and Forecasting using the EM Algorithm. *Journal of Time Series Analysis*, **3** (4), 253–264.
- Snyder, C., T. Bengtsson, P. Bickel, and J. Anderson, 2008: Obstacles to High-Dimensional Particle Filtering. *Monthly Weather Review*, **136** (12), 4629–4640.
- Solonen, A., J. Hakkarainen, A. Ilin, M. Abbas, and A. Bibov, 2014: Estimating model error covariance matrix parameters in extended Kalman filtering. *Nonlinear Processes in Geophysics*, **21** (5), 919–927.
- Stroud, J. R., and T. Bengtsson, 2007: Sequential State and Variance Estimation within the Ensemble Kalman Filter. *Monthly Weather Review*, **135** (9), 3194–3208.

- Stroud, J. R., M. Katzfuss, and C. K. Wikle, 2018: A Bayesian adaptive ensemble Kalman filter for sequential state and parameter estimation. *Monthly Weather Review*, **146** (1), 373–386.
- Tandeo, P., M. Pulido, and F. Lott, 2015: Offline parameter estimation using EnKF and maximum likelihood error covariance estimates: Application to a subgrid-scale orography parametrization. *Quarterly Journal of the Royal Meteorological Society*, **141** (687), 383–395.
- Todding, R., 2015: A lag-1 smoother approach to system-error estimation: sequential method. *Quarterly Journal of the Royal Meteorological Society*, **141** (690), 1502–1513.
- Tremolet, Y., 2007: Model-error estimation in 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, **133** (626), 1267–1280.
- Ueno, G., T. Higuchi, T. Kagimoto, and N. Hirose, 2010: Maximum likelihood estimation of error covariances in ensemble-based filters and its application to a coupled atmosphere-ocean model. *Quarterly Journal of the Royal Meteorological Society*, **136** (650), 1316–1343.
- Ueno, G., and N. Nakamura, 2014: Iterative algorithm for maximum-likelihood estimation of the observation-error covariance matrix for ensemble-based filters. *Quarterly Journal of the Royal Meteorological Society*, **140** (678), 295–315.
- Ueno, G., and N. Nakamura, 2016: Bayesian estimation of the observation-error covariance matrix in ensemble-based filters. *Quarterly Journal of the Royal Meteorological Society*, **142** (698), 2055–2080.
- van Leeuwen, P. J., 2009: Particle filtering in geophysical systems. *Monthly Weather Review*, **137**, 4089–4114.
- Wahba, G., D. R. Johnson, F. Gao, and J. Gong, 1995: Adaptive tuning of numerical weather prediction models: Randomized GCV in three- and four-dimensional data assimilation. *Monthly Weather Review*, **123** (11), 3358–3370.
- Wang, X., and C. H. Bishop, 2003: A Comparison of Breeding and Ensemble Transform Kalman Filter Ensemble Forecast Schemes. *Journal of the Atmospheric Sciences*, **60** (9), 1140–1158.
- Whitaker, J. S., and T. M. Hamill, 2012: Evaluating Methods to Account for System Errors in Ensemble Data Assimilation. *Monthly Weather Review*, **140** (9), 3078–3089.
- Whitaker, J. S., T. M. Hamill, X. Wei, Y. Song, and Z. Toth, 2008: Ensemble Data Assimilation with the NCEP Global Forecast System. *Monthly Weather Review*, **136** (2), 463–482.
- Winiarek, V., M. Bocquet, N. Duhanyan, Y. Rouston, O. Saunier, and A. Mathieu, 2014: Estimation of the caesium-137 source term from the Fukushima Daiichi nuclear power plant using a consistent joint assimilation of air concentration and deposition observations. *Atmospheric Environment*, **82**, 268–279.
- Winiarek, V., M. Bocquet, O. Saunier, and A. Mathieu, 2012: Estimation of errors in the inverse modeling of accidental release of atmospheric pollutant: Application to the reconstruction of the cesium-137 and iodine-131 source terms from the Fukushima Daiichi power plant. *Journal of Geophysical Research: Atmospheres*, **117** (D5).
- Wu, C. F. J., 1983: On the convergence properties of the EM algorithm. *Annals of Statistics*, **11** (1), 95–103.
- Wu, L., M. Bocquet, F. Chevallier, T. Lauvaux, and K. Davis, 2013: Hyperparameter estimation for uncertainty quantification in mesoscale carbon dioxide inversions. *Tellus, Series B: Chemical and Physical Meteorology*, **65** (1), 20 894.
- Ying, Y., and F. Zhang, 2015: An adaptive covariance relaxation method for ensemble data assimilation. *Quarterly Journal of the Royal Meteorological Society*, **141** (692), 2898–2906.
- Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of Initial Estimate and Observation Availability on Convective-Scale Data Assimilation with an Ensemble Kalman Filter. *Monthly Weather Review*, **132** (5), 1238–1253.
- Zhen, Y., and J. Harlim, 2015: Adaptive error covariances estimation methods for ensemble Kalman filters. *Journal of Computational Physics*, **294**, 619–638.
- Zheng, X., 2009: An adaptive estimation of forecast error covariance parameters for Kalman filtering data assimilation. *Advances in Atmospheric Sciences*, **26** (1), 154–160.
- Zhu, M., P. J. van Leeuwen, and J. Amezcua, 2016: Implicit equal-weights particle filter. *Quarterly Journal of the Royal Meteorological Society*, **142** (698), 1904–1919.
- Zhu, M., P. J. van Leeuwen, and W. Zhang, 2017: Estimating model error covariances using particle filters. *Quarterly Journal of the Royal Meteorological Society*, doi:10.1002/qj.3132.
- Zupanski, D., 1997: A General Weak Constraint Applicable to Operational 4DVAR Data Assimilation Systems. *Monthly Weather Review*, **125** (9), 2274–2292.