



Estimating model error covariance matrix parameters in extended Kalman filtering

A. Solonen^{1,2}, J. Hakkarainen², A. Ilin³, M. Abbas³, and A. Bibov¹

¹Lappeenranta University of Technology, Lappeenranta, Finland

²Finnish Meteorological Institute, Helsinki, Finland

³Aalto University, Helsinki, Finland

Correspondence to: A. Solonen (antti.solonen@gmail.com)

Received: 8 August 2013 – Revised: 19 May 2014 – Accepted: 29 July 2014 – Published: 1 September 2014

Abstract. The extended Kalman filter (EKF) is a popular state estimation method for nonlinear dynamical models. The model error covariance matrix is often seen as a tuning parameter in EKF, which is often simply postulated by the user. In this paper, we study the filter likelihood technique for estimating the parameters of the model error covariance matrix. The approach is based on computing the likelihood of the covariance matrix parameters using the filtering output. We show that (a) the importance of the model error covariance matrix calibration depends on the quality of the observations, and that (b) the estimation approach yields a well-tuned EKF in terms of the accuracy of the state estimates and model predictions. For our numerical experiments, we use the two-layer quasi-geostrophic model that is often used as a benchmark model for numerical weather prediction.

1 Introduction

In state estimation, or data assimilation, the goal is to estimate the dynamically changing state of the model, given incomplete and noisy observations. The estimation is usually carried out sequentially: the model prediction made from the previous state is updated with the new observations that become available. State estimation methods need a description of the error that the forward model makes in an assimilation step: otherwise the erroneous model prediction is overweighted when it is combined with new observations, potentially leading to divergence of the method. From the perspective of data assimilation, the model error representation can be viewed as a tunable quantity that has an effect on the performance of the method.

The extended Kalman filter (EKF) is a popular nonlinear data assimilation method. It is a nonlinear extension of the Kalman filter (KF; Kalman, 1960), where the forward model and observation operators are linear, and the model and observation errors are assumed to be additive and normally distributed, which yields direct matrix formulas for updating the model state with observations. In EKF, the nonlinear forward model and the observation operator are linearized, and the KF formulas are applied. The goal in this paper is to study a technique for estimating a parameterized version of a model error covariance matrix.

The model error tuning problem can be viewed as a parameter estimation problem in state space models. One way to estimate static parameters in dynamical state space models is to compute the likelihood of the parameters by “integrating out” the uncertain model state using a filtering technique such as EKF. This “filter likelihood” technique is a well-known tool for parameter estimation in stochastic differential equation (SDE) models (Singer, 2002) and time series analysis (Durbin and Koopman, 2001). In Hakkarainen et al. (2012), the approach was used to estimate parameters of chaotic models. As noted in Hakkarainen et al. (2012), the same approach can be used for estimating the parameters of a model error covariance matrix. In this paper, we study this possibility further. The technique needs a parametric representation of the model error covariance matrix, which can range from something very simple (e.g., diagonal) to complicated representations that take into account, for instance, spatial correlation structures.

The presented approach can be thought of as a generalization of the online error covariance parameter estimation method of Dee (1995), where the model error covariance

matrix parameters are estimated at each assimilation step using a single batch of observations that become available at that step. In the approach presented here, data from several assimilation steps are gathered in the likelihood.

We test the method with the 1600-dimensional two-layer quasi-geostrophic (QG) benchmark model (Pedlosky, 1987) with synthetic data. We show that (a) the importance of the model error covariance matrix calibration depends on the quality of the observations, and that (b) the estimation approach yields a well-tuned EKF in terms of root mean squared (rms) errors of the state estimates and model predictions.

In such a synthetic case, the truth is known, so the true model error can be studied by computing differences between the truth and the forecast model. However, we observe that the true model error is not optimal with respect to the performance of the filter. This happens because filtering methods make certain assumptions and approximations, and the effect of these can be compensated for by appropriately choosing the model error term in the filter. This issue is discussed further in Hakkarainen et al. (2013).

2 Likelihood via filtering

We start this section by introducing how the parameter likelihood can be computed via filtering methods. We introduce briefly the general formulas first, and then proceed to the specific case when EKF is applied. For more details about state estimation theory in the general setting, and about parameter estimation within state space models, refer to, e.g., Särkkä (2013).

Let us consider the following general state space model at time step k with unknown parameters θ :

$$\mathbf{x}_k \sim p(\mathbf{x}_k | \mathbf{x}_{k-1}, \theta) \quad (1)$$

$$y_k \sim p(y_k | \mathbf{x}_k) \quad (2)$$

$$\theta \sim p(\theta). \quad (3)$$

In addition to the unknown dynamically changing model state \mathbf{x}_k , we thus have unknown static parameters θ , from which we might have some prior information $p(\theta)$. The goal in parameter estimation, in Bayesian terms, is to find the posterior distribution $p(\theta | \mathbf{y}_{1:n})$ of the parameters, given a fixed data set $\mathbf{y}_{1:n}$. According to the Bayes formula, the posterior is proportional to the product of the likelihood and the prior: $p(\theta | \mathbf{y}_{1:n}) \propto p(\mathbf{y}_{1:n} | \theta) p(\theta)$. Here, the notation $\mathbf{y}_{1:n} = \{y_1, \dots, y_n\}$ means all observations for n time steps. In the prior term $p(\theta)$, we can include things that we know about θ before collecting any data, such as physical bounds for the parameter values. Here, we assume that the prior is given, and concentrate on the computation of the likelihood $p(\mathbf{y}_{1:n} | \theta)$, which is nontrivial in the case of a state space model.

The general state space model notation given above can be somewhat unfamiliar to readers in the atmospheric sciences

community, and some clarification may be useful. The first equation basically contains the probabilistic model for propagating the state forward: it gives the probability density for the state \mathbf{x}_k , given the value for the previous state \mathbf{x}_{k-1} and the model parameters θ . The second equation contains the observation model; it gives the probability density for observing the value y_k , given a value for the current state \mathbf{x}_k . Presentation of filtering theory often starts from some assumptions of the forms of these densities (such as Gaussian). For the purpose of parameter estimation, it is instructive to develop the likelihood first in a general state space model setup.

Filtering methods (particle filters, Kalman filters, etc.) estimate the dynamically changing model state sequentially. They give the marginal distribution of the state, given the measurements obtained until the current time k . For a given value for θ , filtering methods thus target $p(\mathbf{x}_k | \mathbf{y}_{1:k}, \theta)$. Filters work by iterating two steps: prediction and update. In the prediction step, the current distribution of the state is evolved with the dynamical model to the next time step. In the general notation, the predictive distribution is given by the integral

$$p(\mathbf{x}_k | \mathbf{y}_{1:k-1}, \theta) = \int p(\mathbf{x}_k | \mathbf{x}_{k-1}, \theta) p(\mathbf{x}_{k-1} | \mathbf{y}_{1:k-1}, \theta) d\mathbf{x}_{k-1}, \quad (4)$$

which is known as the Chapman–Kolmogorov equation. When the new observation y_k is obtained, the model state is updated using the Bayes rule with the predictive distribution $p(\mathbf{x}_k | \mathbf{y}_{1:k-1}, \theta)$ as the prior:

$$p(\mathbf{x}_k | \mathbf{y}_{1:k}, \theta) \propto p(y_k | \mathbf{x}_k, \theta) p(\mathbf{x}_k | \mathbf{y}_{1:k-1}, \theta). \quad (5)$$

This posterior is used inside the integral (Eq. 4) to obtain the prior for the next time step.

Using the marginal state posteriors obtained in the filtering method, it is also possible to compute the predictive distribution of the next observation. For observation y_k , the predictive distribution, given all previous observations, can be written as

$$p(y_k | \mathbf{y}_{1:k-1}, \theta) = \int p(y_k | \mathbf{x}_k, \theta) p(\mathbf{x}_k | \mathbf{y}_{1:k-1}, \theta) d\mathbf{x}_k. \quad (6)$$

The term $p(\mathbf{x}_k | \mathbf{y}_{1:k-1}, \theta)$ in the integral is the predictive distribution given by Eq. (4).

Let us now proceed to the original task of estimating static parameters θ from observations $\mathbf{y}_{1:n}$, i.e., computing the posterior distribution $p(\theta | \mathbf{y}_{1:n}) \propto p(\mathbf{y}_{1:n} | \theta) p(\theta)$. Applying the chain rule for joint probability, we obtain

$$p(\mathbf{y}_{1:n} | \theta) = p(y_1 | \theta) \prod_{k=2}^n p(y_k | \mathbf{y}_{1:k-1}, \theta). \quad (7)$$

The likelihood of the whole data set $\mathbf{y}_{1:n}$ can thus be calculated as the product of the predictive distributions of the

individual observations. In the filtering context, the predictive distributions $p(\mathbf{y}_k | \mathbf{y}_{1:k-1}, \boldsymbol{\theta})$ are calculated based on the marginal posterior of the states; see Eq. (6).

The integrals required to construct the likelihood above are usually computationally intractable. In this paper, we use EKF as the filtering method to compute an approximation of the likelihood. We thus now write the state space model in a more familiar form:

$$\mathbf{x}_k = \mathcal{M}(\mathbf{x}_{k-1}) + \mathbf{E}_k(\boldsymbol{\theta}) \tag{8}$$

$$\mathbf{y}_k = \mathcal{K}(\mathbf{x}_k) + \mathbf{e}_k, \tag{9}$$

where \mathcal{M} is the forward model and \mathcal{K} is the observation operator. Note that the unknown parameters $\boldsymbol{\theta}$ now appear in the model error \mathbf{E}_k . In Kalman filtering, it is assumed that model and observation errors are zero mean Gaussian, and that the state and the model error are uncorrelated. Let us assume that the covariance matrix of the model error is parameterized – $\mathbf{E}_k(\boldsymbol{\theta}) \sim N(\mathbf{0}, \mathbf{Q}_k(\boldsymbol{\theta}))$ – and that the observation error covariance matrix is known – $\mathbf{e}_k \sim N(\mathbf{0}, \mathbf{R}_k)$. Now we can apply EKF, and we get the following expression for the predictive distributions needed in the filter likelihood Eq. (7):

$$\mathbf{y}_k | \mathbf{y}_{1:k-1}, \boldsymbol{\theta} \sim N(\mathcal{K}(\mathbf{x}_k^p), \mathbf{C}_k^y), \tag{10}$$

where \mathbf{x}_k^p is the predicted state from the previous time and $\mathbf{C}_k^y = \mathbf{K}_k \mathbf{C}_k^p \mathbf{K}_k^T + \mathbf{R}_k$ is the covariance matrix of the predictive distribution, containing both the model prediction covariance matrix \mathbf{C}_k^p and the observation error covariance matrix \mathbf{R}_k . The matrix \mathbf{K}_k is the linearized observation model. In EKF, the prediction covariance matrix is computed as $\mathbf{C}_k^p = \mathbf{M}_k \mathbf{C}_{k-1}^{\text{est}} \mathbf{M}_k^T + \mathbf{Q}_k(\boldsymbol{\theta})$, where \mathbf{M}_k is the linearized forward model, $\mathbf{C}_{k-1}^{\text{est}}$ is the covariance estimate of the previous time step and $\mathbf{Q}_k(\boldsymbol{\theta})$ is the parameterized model error covariance matrix.

Now, applying the general formula (Eq. 7) to the EKF case, the likelihood of observing data $\mathbf{y}_{1:n}$ given parameters $\boldsymbol{\theta}$ can be written as

$$p(\mathbf{y}_{1:n} | \boldsymbol{\theta}) \propto \exp\left(-\frac{1}{2} \sum_{k=1}^n \left[\mathbf{r}_k^T (\mathbf{C}_k^y)^{-1} \mathbf{r}_k + \log |\mathbf{C}_k^y| \right]\right), \tag{11}$$

where $\mathbf{r}_k = \mathbf{y}_k - \mathcal{K}(\mathbf{x}_k^p)$ are the prediction residuals and $|\cdot|$ denotes the matrix determinant. The normalization “constants” of the likelihood terms depend on $\boldsymbol{\theta}$ through the covariances \mathbf{C}_k^p , and the determinant term therefore needs to be included. Note that the above likelihood is only an approximation to the true likelihood (Eq. 7), and the accuracy of this approximation depends on how well the EKF assumptions (linear model used in error propagation, model error assumed independent of the state) are met. In practice, using the EKF likelihood in parameter estimation often yields good results; see, for instance, Singer (2002), Hakkarainen et al. (2012), Mbalawata et al. (2013) and the numerical examples of this paper.

In Dee (1995), model error covariance parameters were estimated online at each assimilation time step, using only the observations that become available at that specific step. In our notation, this would correspond to having just one term in the exponent of Eq. (11) instead of a sum; that is, the approach presented here can be thought of as a generalization of the approach in Dee (1995), where data from several assimilation steps are gathered in the likelihood.

Note that we could also include some parameters in the forward model, and we would have $\mathcal{M}(\mathbf{x}_k, \boldsymbol{\theta})$ instead of $\mathcal{M}(\mathbf{x}_k)$. In this paper, we focus only on the model error parameters, but the same technique also applies to model parameters, as demonstrated in Hakkarainen et al. (2012). In principle, we could also assume a model error with a non-zero mean and estimate the parameterized mean of the model error as well, and possibly correct for systematic bias in the model, but this possibility is not pursued further here.

This method assumes that there is a parametric representation of the model error covariance $\mathbf{Q}_k(\boldsymbol{\theta})$ available. In the examples presented in this paper, the model error is assumed to be static over time; we have $\mathbf{Q}_k(\boldsymbol{\theta}) = \mathbf{Q}(\boldsymbol{\theta})$ for all time steps k .

3 Numerical experiments with the two-layer quasi-geostrophic model

3.1 Model description

The two-layer quasi-geostrophic model simulates atmospheric flow for the geostrophic (slow) wind motions. This model can be used as a benchmark for data assimilation in numerical weather prediction (NWP) systems, as it supports some features common to operational weather models, such as baroclinic instability. At the same time, the QG model has a relatively low computational complexity, and requires no special hardware to run. The geometrical domain of the model is specified by a cylindrical surface vertically divided into two “atmospheric” layers that can interact through the interface between them. The model also accounts for an orographic component that defines the surface irregularities affecting the bottom layer of the model. When the geometrical layout of the two-layer QG model is mapped onto a plane, it appears as shown in Fig. 1. In the figure, parameters U_1 and U_2 denote mean zonal flows in the top and the bottom atmospheric layers, respectively.

The model operates in terms of potential vorticity and stream function, where the latter one is analogous to pressure. The assumption of quasi-geostrophic motion leads to a coupled system of partial differential equations (PDEs) describing a conservation law for potential vorticity, given as

$$\frac{\mathcal{D}_1 q_1}{\mathcal{D}t} = 0, \quad \frac{\mathcal{D}_2 q_2}{\mathcal{D}t} = 0, \tag{12}$$

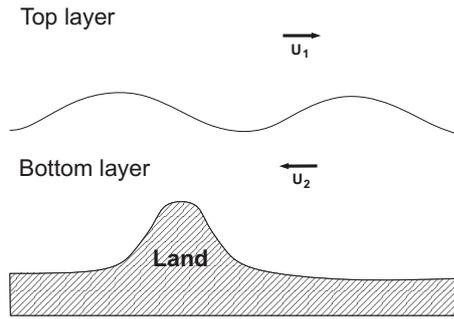


Figure 1. Geometrical layout of the two-layer quasi-geostrophic model.

where D_i denote the substantial derivatives for latitudinal wind u_i and longitudinal wind v_i , defined as $\frac{D_i}{Dt} = \frac{\partial}{\partial t} + u_i \frac{\partial}{\partial x} + v_i \frac{\partial}{\partial y}$, q_i denote the potential vorticity functions, and index i specifies the top atmospheric layer ($i = 1$) and the bottom layer ($i = 2$). Interaction between the layers, as well as the relation between the potential vorticity q_i and the stream function ψ_i , is modeled by the following system of PDEs:

$$q_1 = \nabla^2 \psi_1 - F_1 (\psi_1 - \psi_2) + \beta y, \quad (13)$$

$$q_2 = \nabla^2 \psi_2 - F_2 (\psi_2 - \psi_1) + \beta y + R_s. \quad (14)$$

Here, R_s and β denote the dimensionless orography component and the northward gradient of the Coriolis parameter, which we hereafter denote as f_0 . The relations between the model physical attributes and dimensionless parameters that appear in Eqs. (13)–(14) are as follows:

$$F_1 = \frac{f_0^2 L^2}{g' D_1}, \quad F_2 = \frac{f_0^2 L^2}{g' D_2}, \quad g' = g \frac{\Delta \theta}{\bar{\theta}},$$

$$R_s = \frac{S(x, y)}{\eta D_2}, \quad \beta = \beta_0 \frac{L}{U},$$

where D_1 and D_2 are the layer depths, $\Delta \theta$ defines the potential temperature change across the layer interface, $\bar{\theta}$ is the mean potential temperature, g is the acceleration of gravity, $\eta = \frac{U}{f_0 L}$ is the Rossby number associated with the defined system, and $S(x, y)$ and β_0 are dimensional representations of $R_s(x, y)$ and β , respectively.

The system of Eqs. (12)–(14) defines the two-layer quasi-geostrophic model. The state of the model, and thus the target of the estimation, is the stream function ψ_i . For the numerical solution of the system, we consider potential vorticity functions q_1 and q_2 to be known, and invert the spatial Eqs. (13) and (14) for ψ_i . More precisely, we apply ∇^2 to Eq. (13), and subtract F_1 times (Eq. 14) and F_2 times (Eq. 13) from the result, which yields the following equation:

$$\nabla^2 \left[\nabla^2 \psi_1 \right] - (F_1 + F_2) \left[\nabla^2 \psi_1 \right]$$

$$= \nabla^2 q_1 - F_2 (q_1 - \beta y) - F_1 (q_2 - \beta y - R_s). \quad (15)$$

Equation (15) can be treated as a non-homogeneous Helmholtz equation with negative parameter $-(F_1 + F_2)$ and unknown $\nabla^2 \psi_1$. Once $\nabla^2 \psi_1$ is solved, the stream function for the top atmospheric layer is determined by a Poisson equation. The stream function for the bottom layer can be found by plugging the obtained value for ψ_1 into Eqs. (13) and (14) and by solving the equations for ψ_2 . The potential vorticity functions q_i are evolved over time by a numerical advection procedure that models the conservation Eq. (12). For more details on the QG model, refer to Fisher et al. (2011).

3.2 Experiment setup and results

We study the model error parameter estimation problem with the QG model described above. In our experiments, we run the model with a 20×40 grid in each layer, and the dimension of the state vector is thus 1600. To generate data, we run the model with a 1 h time step with layer depths $D_1 = 6000$ and $D_2 = 4000$. Data is generated at every 6th step (the assimilation step is thus 6 h) by adding random noise with a given standard deviation σ_y to a given number of randomly chosen grid points. For the EKF estimation, bias is introduced to the forward model by using the wrong layer depths $\tilde{D}_1 = 5500$ and $\tilde{D}_2 = 4500$. To illustrate the model and the observations, a snapshot of a single step of the EKF estimation is given in Fig. 2.

We apply two different parameterizations for $\mathbf{Q}(\theta)$, a simple diagonal parameterization and a more complicated one that includes horizontal and vertical correlations. First, we simply study how important the model error term is in terms of EKF accuracy, with various observation settings. We then test the filter likelihood computation for the two selected $\mathbf{Q}(\theta)$ matrices. As a validation metric, we use the rms error of the state estimates and the model predictions. The likelihood values and the validation metrics are computed using separate “training data” and validation data.

The filter likelihood approach attempts to find a $\mathbf{Q}(\theta)$ so that the model predictions fit the observations with the correct accuracy (forecast error + measurement error), and we therefore expect this approach to yield reasonably good forecast error estimates as well, provided that the EKF assumptions are met. In Solonen and Järvinen (2013), a similar estimation technique was used to estimate the parameters of a small-scale ensemble prediction system (EPS), and there the approach produced a good representation of the forecast uncertainty. In order to verify the realism of the forecast error covariance matrix in this setup, we compare the squared mean variance of the forecast error covariance matrix against the true rms forecast error.

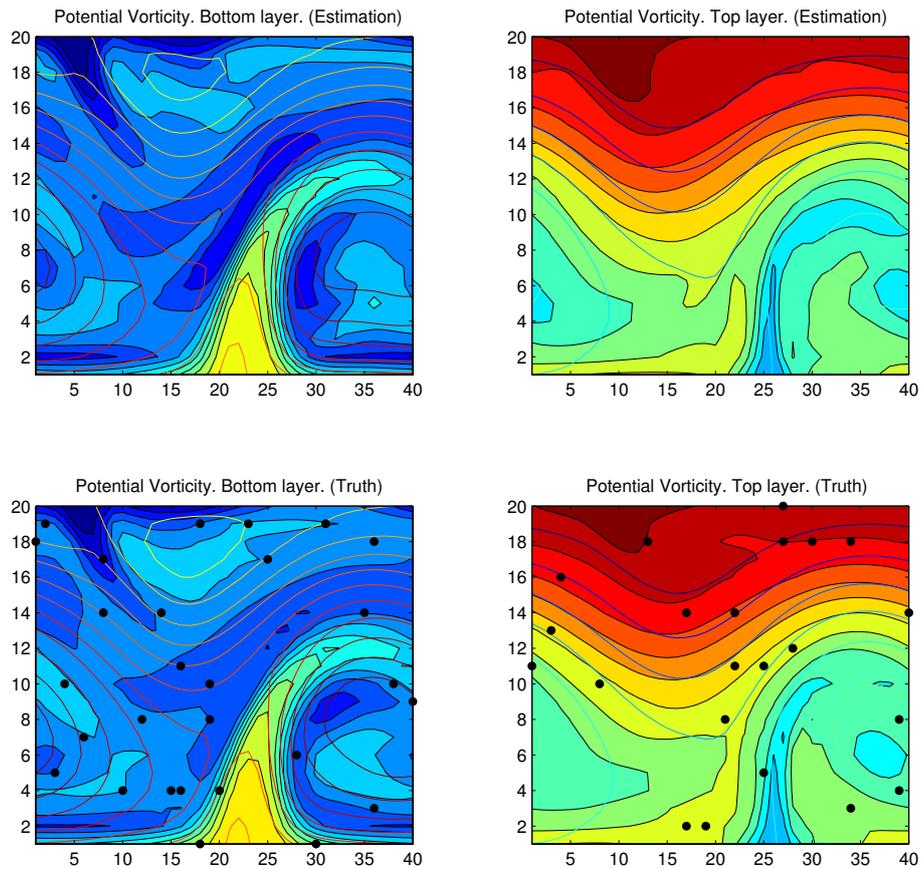


Figure 2. The true state (top-row panels) and the EKF estimate (bottom-row panels) for the bottom layer (left-column panels) and for the top layer (right-column panels) in a 20×40 grid for each layer using 50 randomly chosen observation locations (black dots). Filled contours describe the potential vorticity, and line contours describe the stream function.

3.2.1 The effect of observations to model error calibration

First, we study the relevance of the model error covariance matrix calibration by running EKF with various observation settings and various levels for the model error. We vary the number of observations used for the estimation at each assimilation step (20, 50 and 100), and the observation error standard deviation ($\sigma_y = 0.1$ and $\sigma_y = 1$). As the model error covariance matrix, we use the simplest possible parameterization, $\mathbf{Q} = \lambda \mathbf{I}$. For each observation setting, we run EKF with different values for the model error variance λ .

The results are shown in Fig. 3. One can see that with a large enough number of accurate enough observations (left panel, red curve), the model error calibration has very little effect on the accuracy of the filter; λ can vary many orders of magnitude without any significant difference in the rms errors of the state estimates. Reducing the number of observations (left panel, black and green curves) makes the calibration of the model error slightly more meaningful, and there seems to be an optimum value for λ that yields the most accurate EKF. Still, one can see that the benefit of optimizing λ

is limited in this case; with large values for λ , EKF still converges, and the state estimates are almost as accurate as with the optimized λ .

When the observation error standard deviation is increased from $\sigma_y = 0.1$ to $\sigma_y = 1$ (right panel), the situation changes. Now the model error variance has a clearer impact on the accuracy of the filter, and substantial improvements in the filter can be achieved by correctly choosing the model error covariance parameters.

We conclude that the relevance of the model error calibration depends on the quality of the observations. If we have a large number of accurate observations available, the model error might not matter much. On the other hand, if the information of the observations is limited, the model error has to be tuned accurately to make the filter work properly. From the Bayesian perspective, this result is natural: if the observations do not identify the state properly, the prior has to be tuned carefully to make the estimation work.

3.2.2 Simple model error covariance matrix

To test the likelihood calculation, we first test the simple diagonal parameterization $\mathbf{Q} = \lambda \mathbf{I}$. We compute the likelihood from a training period of 100 assimilation steps, and compute the rms errors of the state estimates using a separate validation period of 100 steps. At each assimilation step, we use 100 observations with noise standard deviation $\sigma_y = 1$.

In Fig. 4, we plot the negative log-likelihood values of the training period to rms errors of the validation period. There is a clear correlation between the likelihood and the validation rms error, maximizing the likelihood results in optimal filter accuracy.

3.2.3 More complicated model error covariance matrix

Next, we perform the same experiment as above, but use a slightly more complicated covariance matrix parameterization. We use a Gaussian covariance function with three parameters, and for each layer, we define the covariance matrix elements as

$$\mathbf{Q}_{ij} = \begin{cases} \tau^2 + \sigma^2 \exp\left(-\frac{d(x_i, x_j)^2}{2\alpha^2}\right) & \text{when } i = j \\ \sigma^2 \exp\left(-\frac{d(x_i, x_j)^2}{2\alpha^2}\right) & \text{when } i \neq j, \end{cases} \quad (16)$$

where $d(x_i, x_j)$ denotes the distance between two grid points, $\sigma^2 > 0$ is the variance parameter, $\alpha > 0$ is the correlation length, and $\tau^2 > 0$ is a small positive nugget term that ensures that the matrix is positive definite. In addition, we estimate the vertical correlation $\rho \in [0, 1]$ between the two layers. We thus have four parameters altogether: $\theta = (\tau^2, \sigma^2, \alpha, \rho)$.

We test randomly chosen parameter values uniformly in the intervals $\sigma^2 \in [0, 0.05]$, $\alpha \in [0, 10]$, $\rho \in [0, 1]$ and $\tau^2 \in [0, 0.01]$. For each parameter combination, we compute the likelihood values and rms errors for validation. The results are shown in Fig. 5 (left panel). Again, we see a clear correlation between the likelihood values and the rms errors: maximizing the likelihood results in an accurate EKF. To validate the results further, we compute the rms errors of forecasts launched from the EKF state estimates with different model error covariance matrix parameters. The results are shown in Fig. 5 (right panel). One can see that the parameters with a high likelihood also validate well in terms of forecast skill.

In a synthetic case, such as here, we know the truth behind the observations, and we can compute “samples” of the model error by running the truth model and the biased forecast model, starting from the same initial value and collecting the differences. This allows us to estimate the “true” model error covariance matrix. We estimated the covariance from 2000 samples of the error in two ways: first, we computed the full matrix directly with the empirical covariance estimation formula, and then estimated the parameters of the covariance

function (Eq. 16) based on the samples. We plugged these matrices into EKF and computed the rms errors for the validation period. The results obtained in this way are shown in the left part of Fig. 5. Surprisingly, these matrices do not validate that well in terms of filter accuracy. We believe that the reason is that EKF is an approximative filter – it uses linearizations and assumes, for instance, that the model error and state are independent – and the imperfections in the method can to some extent be compensated for by calibrating the model error covariance matrix.

3.2.4 Verification of the forecast error covariance matrix

In order to verify the quality of the Kalman filter forecast error covariance matrix $\mathbf{C}_k^p = \mathbf{M}_k \mathbf{C}_{k-1}^{\text{est}} \mathbf{M}_k^T + \mathbf{Q}_k(\theta)$, the following two metrics are considered:

$$m_1(k) = \sqrt{(\mathbf{x}_k^p - \mathbf{x}_k^{\text{true}})^2}, \quad (17)$$

$$m_2(k) = \sqrt{(\sigma_k^p)^2}. \quad (18)$$

The first metric is the true rms forecast error, where the forecast \mathbf{x}_k^p is calculated from the previous Kalman filter estimate. The second metric is the squared mean variance of the forecast error covariance matrix. The mean in both cases is calculated over the 1600-dimensional state space.

In Fig. 6, we plot these two metrics using five different parameter combinations and the “true” model error covariance matrix (obtained via samples of the model error; see Sect. 3.2.3). For the parameter combinations, we selected the points that give the best and worst cost function values, and the points that correspond to the three quartile points of the cost function values (indicated by Q1, median and Q3 in Fig. 6).

From Fig. 6, we can observe that, using the best cost function point, the two metrics give a similar mean error level, showing that – on average – the Kalman filter forecast error (co)variance is realistic. This observation is also valid for the other points for which the cost function value is close to a minimum (grey lines in Fig. 6). For the other parameter combinations, we observe that the estimated and true forecast errors do not match: the forecast error is overestimated, and the difference grows gradually when going towards the worst parameter combination. The “true” model error covariance matrix, on the other hand, underestimates the forecast (co)variance, which is anticipated, as it does not take into account the imperfections of the EKF method discussed earlier.

The validation here was done using the more complicated model error parameterization (Eq. 16). We note that if metrics m_1 and m_2 are calculated using the simple diagonal covariance matrix $\mathbf{Q} = \lambda \mathbf{I}$, too low (high) λ values give on average too low (high) forecast error (co)variance, as expected. Near the optimum, the level of the forecast error covariance matrix is realistic (not shown).

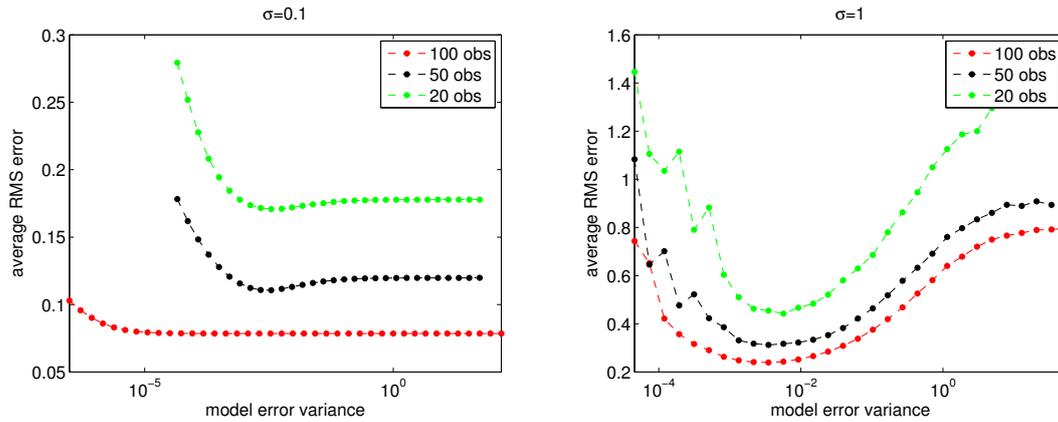


Figure 3. Model error variances λ vs. average rms errors of the state estimates with a varying number of observations and observation errors $\sigma_y = 0.1$ (left panel) and $\sigma_y = 1$ (right panel).

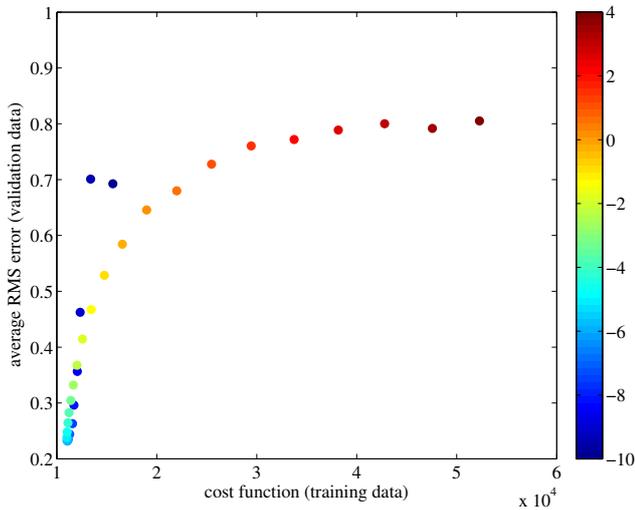


Figure 4. Negative log likelihood of the training data vs. the average rms error of the filter with the validation data. Colors indicate the logarithm of the model error variance λ .

4 Discussion and conclusions

In this paper, we consider the problem of calibrating the model error covariance matrix \mathbf{Q} in extended Kalman filtering (EKF). The matrix \mathbf{Q} is commonly seen as a tuning parameter in EKF, and is often postulated by the user in an ad hoc manner. We study a technique for objectively estimating the parameters θ of a parametric version of the matrix, $\mathbf{Q}(\theta)$, based on indirect and noisy observations of the model state. The approach is based on approximating the likelihood of the parameters θ using the EKF output. This “filter likelihood” method is tested with the two-layer quasi-geostrophic model that is often used as a benchmark case in numerical weather prediction studies.

One of our findings is that the relevance of the calibration of \mathbf{Q} depends on the quality of the observations. The less

information the observations contain about the model state, the more carefully the prior, a part of which \mathbf{Q} is, needs to be tuned. On the other hand, if we have enough accurate observations, accurate optimization of \mathbf{Q} might not be that beneficial. Secondly, we conclude that the filter likelihood approach works well in our test cases; maximizing the likelihood results in accurate EKF in terms of the rms errors of the state estimates and model predictions. In addition, the points that give a high likelihood validate well in terms of the quality of the forecast error estimates.

Our experiments in this paper are synthetic in the sense that we generate observations with the “true model” and introduce bias into the model that is used for estimation. In such a case, one can estimate the “true” model error by running predictions with the truth model and the biased forecast model, and collecting the differences between the predictions. Our experiments suggest that the model error obtained in this way is not optimal in terms of filter accuracy. A reason might be that the model error can be used to account for approximations and assumptions made in the filtering method. The consequence is that each filtering method should be tuned separately: the \mathbf{Q} that works best in EKF might not be optimal for other filtering methods.

In this paper, the focus was on the extended Kalman filter. However, similar model error parameters appear in many other data assimilation methods as well, like, for instance, in the weak-constraint 4D-Var (Fisher et al., 2005) and ensemble Kalman filters (Evensen, 2007). In many so-called ensemble square root Kalman filters (Tippett et al., 2003), the model error is neglected, but covariance inflation techniques are used to account for the resulting underestimation of the uncertainty. We note that the parameters related to covariance inflation can also be tuned with the presented approach, as demonstrated in Hakkarainen et al. (2013). A problem with some ensemble methods is that they contain random perturbations, which can complicate the optimization process,

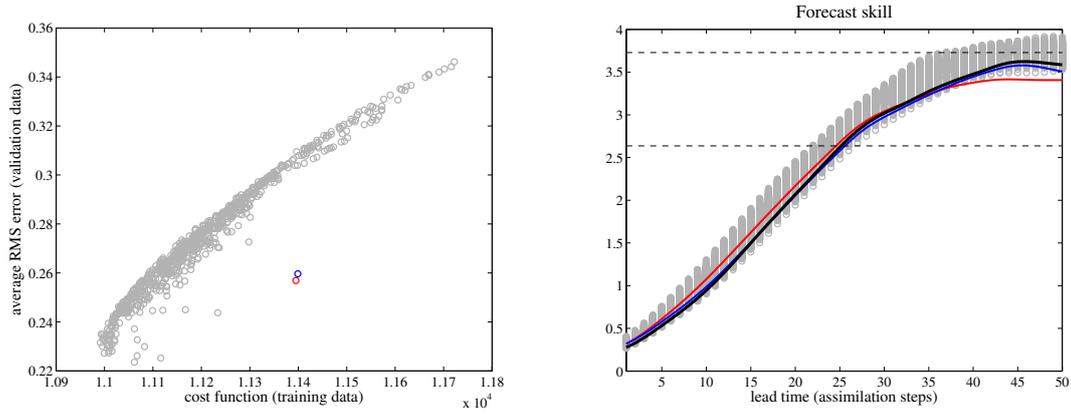


Figure 5. Left panel: negative log likelihood of the training data vs. the average rms error of the filter with the validation data. Right panel: the average forecast rms errors with different covariance matrix parameters. Red and blue colors indicate the results acquired with the full and parametrized “true” model error covariance matrices, respectively. The black line indicates the forecast skill acquired with the parameters that give the smallest negative log likelihood. Dashed vertical lines indicate climatological forecast skill and error saturation level.

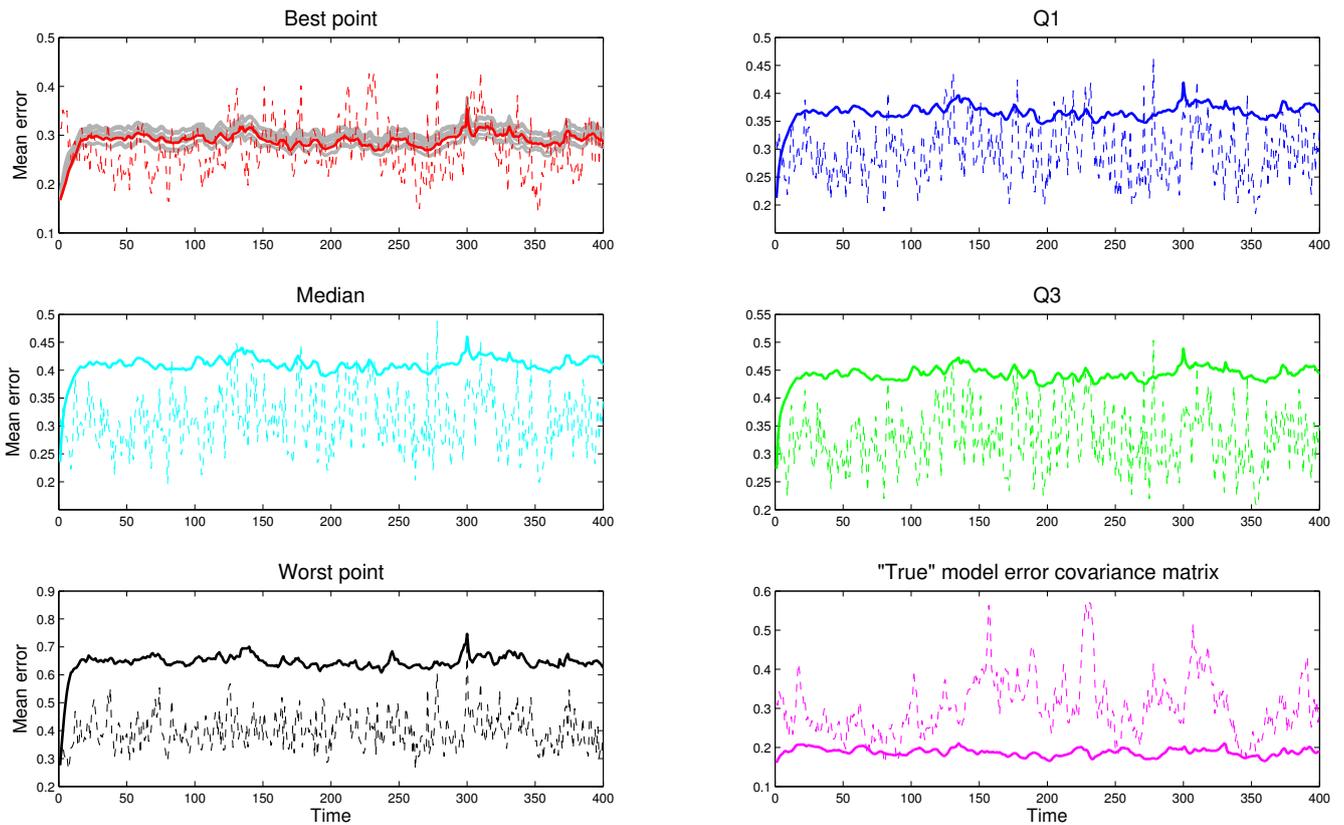


Figure 6. The true forecast error vs. the Kalman filter forecast error, calculated using different parameter combinations, and the “true” model error covariance matrix obtained from samples of model error. Dashed lines indicate the true rms forecast error and solid lines the squared mean variance of the forecast error covariance matrix. In the first subplot, the latter metric is also calculated from points that give the ten lowest cost function values (grey lines). The subplots Q1, median and Q3 give the results for parameter values that correspond to the three quartiles of the cost function values. Note the difference in scale in each subplot.

since the likelihood is stochastic, as noted in Hakkarainen et al. (2012) and Dowd (2011).

Here, we simply confirm that the likelihood approach works for calibrating the model error, and do not consider algorithms for maximizing or exploring the likelihood surface. A suitable method is case dependent. For simple models, standard optimization or, for instance, Markov chain Monte Carlo algorithms, are available. If the model is computationally expensive, one needs an efficient method to explore the surface with as few likelihood evaluations as possible. For instance, methods that apply empirical approximations (or “emulators”) of the likelihood surface seem promising here; see, e.g., Rasmussen (2003). These topics are left for future research.

Acknowledgements. The research has been supported by the Academy of Finland (NOVAC and INQUIRE projects) and the Centre of Excellence in Inverse Problems Research. The authors would like to thank the editor and the two anonymous reviewers for their insightful comments that helped to improve the manuscript.

Edited by: Z. Toth

Reviewed by: two anonymous referees

References

- Dee, D. P.: On-line estimation of error covariance parameters for atmospheric data assimilation, *Mon. Weather Rev.*, 123, 1128–1145, doi:10.1175/1520-0493(1995)123<1128:OLEOEC>2.0.CO;2, 1995.
- Dowd, M.: Estimating parameters for a stochastic dynamic marine ecological system, *Environmetrics*, 22, 501–515, doi:10.1002/env.1083, 2011.
- Durbin, J. and Koopman, S. J.: *Time series analysis by state space methods*, Oxford University Press, New York, 2001.
- Evensen, G.: *Data assimilation: The ensemble Kalman filter*, Springer, Berlin, 2007.
- Fisher, M., Leutbecher, M., and Kelly, G.: On the equivalence between Kalman smoothing and weak-constraint four-dimensional variational data assimilation, *Q. J. Roy. Meteorol. Soc.*, 131, 3235–3246, doi:10.1256/qj.04.142, 2005.
- Fisher, M., Tremolet, Y., Auvinen, H., Tan, D., and Poli, P.: *Weak-constraint and long window 4DVAR*, Technical Report 655, European Centre for Medium-Range Weather Forecasts, Reading, UK, 2011.
- Hakkarainen, J., Ilin, A., Solonen, A., Laine, M., Haario, H., Tamminen, J., Oja, E., and Järvinen, H.: On closure parameter estimation in chaotic systems, *Nonlin. Processes Geophys.*, 19, 127–143, doi:10.5194/npg-19-127-2012, 2012.
- Hakkarainen, J., Solonen, A., Ilin, A., Susiluoto, J., Laine, M., Haario, H., and Järvinen, H.: The dilemma on the uniqueness of weather and climate model closure parameters, *Tellus A*, 65, 20147, doi:10.3402/tellusa.v65i0.20147, 2013.
- Kalman, R. E.: A New Approach to Linear Filtering and Prediction Problems, *J. Basic Eng.*, 82, 35–45, 1960.
- Mbalawata, I., Särkkä, S., and Haario, H.: Parameter estimation in stochastic differential equations with Markov chain Monte Carlo and non-linear Kalman filtering, *Comput. Stat.*, 28, 1195–1223, doi:10.1007/s00180-012-0352-y, 2013.
- Pedlosky, J.: *Geophysical Fluid Dynamics*, 2nd Edn., Springer-Verlag, New York, 1987.
- Rasmussen, C. E.: Gaussian processes to speed up hybrid Monte Carlo for expensive Bayesian integrals, *Bayesian Stat.*, 7, 651–659, 2003.
- Särkkä, S.: *Bayesian Filtering and Smoothing*, Cambridge University Press, 2013.
- Singer, H.: Parameter estimation of nonlinear stochastic differential equations: Simulated maximum likelihood versus extended Kalman filter and Itô–Taylor expansion, *J. Comput. Graph. Stat.*, 11, 972–995, doi:10.1198/106186002808, 2002.
- Solonen, A. and Järvinen, H.: An approach for tuning ensemble prediction systems, *Tellus A*, 65, 20594, doi:10.3402/tellusa.v65i0.20594, 2013.
- Tippett, M., Anderson, J., Bishop, G., Hamill, T., and Whitaker, J.: Ensemble Square Root Filter, *Mon. Weather Rev.*, 131, 1485–1490, doi:10.1175/1520-0493(2003)131<1485:ESRF>2.0.CO;2, 2003.