

# A dilemma of the uniqueness of weather and climate model closure parameters

By JANNE HAKKARAINEN<sup>1,2\*</sup>, ANTTI SOLONEN<sup>1,2</sup>, ALEXANDER ILIN<sup>3</sup>,  
JOUNI SUSILUOTO<sup>1,4</sup>, MARKO LAINE<sup>1</sup>, HEIKKI HAARIO<sup>2</sup> and  
HEIKKI JÄRVINEN<sup>4</sup>, <sup>1</sup>*Finnish Meteorological Institute, Helsinki, Finland;*  
<sup>2</sup>*Lappeenranta University of Technology, Lappeenranta, Finland;* <sup>3</sup>*Aalto University School of Science,*  
*Espoo, Finland;* <sup>4</sup>*University of Helsinki, Helsinki, Finland*

(Manuscript received 21 November 2012; in final form 19 March 2013)

## ABSTRACT

Parameterisation schemes of subgrid-scale physical processes in atmospheric models contain so-called closure parameters. Their precise values are not generally known; thus, they are subject to fine-tuning for achieving optimal model performance. In this article, we show that there is a dilemma concerning the optimal parameter values: an identical prediction model formulation can have two different optimal closure parameter value settings depending on the level of approximations made in the data assimilation component of the prediction system. This result tends to indicate that the prediction model re-tuning in large-scale systems is not only needed when the prediction model undergoes a major change, but also when the data assimilation component is updated. Moreover, we advocate an accurate albeit expensive method based on so-called filter likelihood for the closure parameter estimation that is applicable in fine-tuning of both prediction model and data assimilation system parameters. In this article, we use a modified Lorenz-95 system as a prediction model and extended Kalman filter and ensemble adjustment Kalman filter for data assimilation. With this setup, we can compute the filter likelihood for the chosen parameters using the output of the two versions of the Kalman filter and apply a Markov chain Monte Carlo algorithm to explore the parameter posterior distributions.

*Keywords:* model tuning, Markov chain Monte Carlo, likelihood, filter formulation

## 1. Introduction

Long-term improvements in atmospheric general circulation models (GCMs) used in numerical weather prediction and climate simulations originate from gradually refined representations of atmospheric phenomena, especially of the subgrid-scale physical processes. In these models, the processes are represented by parameterisation schemes where the subgrid-scale variability is expressed using model variables of the resolved scales. Moreover, the schemes contain so-called closure parameters. The purpose of these model parameters is to encapsulate some atmospheric processes or properties which are not affordable for explicit modelling. The development process of the physical parameterisation schemes involves basically two steps: derivation of individual parameterisation schemes, and constraining of the schemes with large-scale observations

(Lohmann et al., 2007). Here, large-scale observations means observations that are available for model tuning at the GCM scale. In the first step, dedicated laboratory measurements or focused field measurements campaigns are used, if a scheme cannot be derived from the first principles. The second step accounts for the fact that individual schemes are often valid for small scales, while in the context of GCMs, the schemes are applied in a relatively coarse resolution over a wide range of spatial and temporal scales. Thus, the schemes of interest have to be tuned such that the model indeed simulates the key atmospheric quantities within the uncertainty of some large-scale observational constraint. In practice, the closure (or, tuning) parameters provide the necessary degrees-of-freedom so that a realistic model response can be obtained.

In GCMs, a comprehensive set of physical parameterisation schemes are assembled together and embedded in the solver of the atmospheric dynamics. Time-space truncation and numerical approximations render, however, all

\*Corresponding author.  
email: janne.hakkarainen@fmi.fi

model components to some extent imperfect. These model imperfections lead to systematic and random simulation errors. The systematic errors appear as a climate drift in the model simulations. The geographical patterns of these errors tend to develop very early in the simulation (Jung, 2005), and the so-called systematic initial tendency errors can be used for diagnosis and attribution of model errors to some particular atmospheric processes (Klinker and Sardeshmukh, 1992). In fact, there are attempts to estimate the initial tendency model errors and thus parameterise unrepresented physical phenomena in order to reduce systematic model errors (Kaas et al., 1999; D’Andrea and Vautard, 2000). Clearly, in the model simulations, there are multitudes of errors with different sources and the errors are in mutual non-linear interaction. Thus, it is not a surprise that it is very hard to improve the general model performance by improving the realism of representation of individual physical parameterisations (Jakob, 2010). Therefore, for a given model formulation, it is natural to strive towards optimal predictive skill by tuning the available degrees-of-freedom, that is, the closure parameter values. The estimation techniques range from trial-and-error to sophisticated parameter estimation techniques (e.g. Severijns and Hazeleger, 2005; Kunz et al., 2008; Järvinen et al., 2010; Neelin et al., 2010). The closure parameter values corresponding to the optimal predictive skill naturally depend on the level of approximation of the prediction system.

There is thus a dilemma on the uniqueness of weather and climate model closure parameters: one can either anchor the parameter values to the observable truth obtained, for example, via measurements from focused field campaigns, or tune the parameter values to account for the imperfections in the prediction system. It is qualitatively obvious that these two do not necessarily coincide and a practical optimum is a compromise between the two. In this article, we demonstrate that this dilemma exists. An outline of the demonstration is as follows. We use a low-order prediction system where synthetic observations are created using a modified version of the Lorenz-95 model (Lorenz, 1995; Wilks, 2005). As a prediction model, we use the standard Lorenz-95 model where the subgrid-scale effects on resolved scales are represented by a linear bulk parameterisation containing two closure parameters. Thus, the prediction model is imperfect. Data assimilation of the synthetic observations is performed using either extended Kalman filter (EKF) (Kalman, 1960) or ensemble adjustment Kalman filter (EAKF) (Anderson, 2001). We observe that the optimal prediction model corresponds to a unique but different parameter setting depending on the choice of the data assimilation component. The conclusion is that the dilemma exists.

Since the optimal closure parameter values depend on the implementation of the prediction system, such as the choice of the data assimilation algorithm, methodology for tuning the parameters specifically for each prediction system is needed. Here, we present such a method, using a combination of the so-called filter likelihood approach, where the likelihood is formulated based on the output of the filtering methods (Singer, 2002; Hakkarainen et al., 2012) and a Markov chain Monte Carlo algorithm (Haario et al., 2001, MCMC). We show that the approach yields parameter values that are close to the ones corresponding to optimal filter accuracy. In addition, we demonstrate how the approach can be implemented via the Data Assimilation Research Testbed software environment (Anderson et al., 2009, DART).

## 2. Methods and experimentation

In this section, an introduction to parameter estimation using the filter likelihood approach is given. It is then tested using the modified Lorenz-95 model.

### 2.1. On state estimation methods

One of the most common techniques for state estimation is Kalman filtering (Kalman, 1960). The basic idea in all filters is to obtain the posterior state estimate  $x_k^{\text{est}}$  and some error statistics, typically the error covariance matrix  $C_k^{\text{est}}$ , given the prior information  $x_k^p$  and the observations  $y_k$  at time step  $k$ .

The posterior state estimate can be transported to the next filtering time step to become the new prior state by using the forecast model. For the state vector, this is usually straightforward, but the fundamental question in filtering is how to transport the uncertainty in time. In EKF, this problem is solved by calculating the prior covariance matrix  $C_k^p$  at time  $k$  as

$$C_k^p = M_k C_{k-1}^{\text{est}} M_k^T + Q_k, \quad (1)$$

where  $Q_k$  is the model error covariance matrix and  $M_k$  is the tangent-linear forecast model. If the dimension of the state space is high, such as in large-scale atmospheric models, eq. (1) becomes impossible to compute without approximations.

One approximative approach is ensemble filtering, where the uncertainty is transported to the next filtering time step by transporting an ensemble of states using the forecast model. The difficulty (and variety) of the ensemble methods lies in how to update the prior ensemble in the analysis step. There are many alternatives, for instance the ensemble adjustment Kalman filter (Anderson, 2001, 2010), which belongs to the family of the so-called ensemble square root

filters (Tippett et al., 2003). In this study, the EAKF was chosen because it is a *deterministic* ensemble based filter, that is, it does not include randomisation of the ensembles. The non-randomness is important, since in Section 2.3 we want to create a likelihood function based on the filtering outputs and the randomness in the likelihood would complicate the parameter inference (Dowd, 2011; Hakkarainen et al., 2012). Also, EAKF was chosen because it is conveniently available in the data assimilation research testbed (Anderson et al., 2009, DART) and widely applied (e.g. Schwartz et al., 2011; Torn and Davis, 2012).

In EKF, the modelling errors of the forecast model are taken into account via the error covariance matrix  $Q_k$ . In EAKF (and in ensemble square root filtering in general), the neglect of unknown model errors can lead to overconfidence in the prior state and hence ignorance of the observations in the analysis, which can lead to filter divergence. This is effectively circumvented by introducing covariance inflation factors (Anderson and Anderson, 1999). In addition, the sampling errors related to the limited ensemble size (undersampling), are mostly removed by localisation (Hamill et al., 2001), where observations affect only nearby grid points.

## 2.2. On parameter estimation

Parameter estimation techniques developed for the atmospheric models can be divided in two categories: *online* and *offline* methods. In online methods, it is assumed that model parameters are not static quantities, but can evolve adaptively, for example, as a part of a state estimation process (e.g. Annan et al., 2005) and are thus computationally relatively cheap to implement. Typically, a statistical interpretation is lacking in online methods, since the rate of change of the parameters is controlled by the user. In this article, these online methods are not discussed further. Instead we use offline methodology, where a predetermined (training) set of observations  $y$  is used for tuning the parameter values.

In Bayesian methodology (e.g. Gelman et al., 2003), the knowledge about the unknown parameters is inferred from the posterior distribution:

$$p(\theta|y) \propto p(\theta)p(y|\theta), \quad (2)$$

which is evaluated using the prior  $p(\theta)$  and the likelihood  $p(y|\theta)$ . The prior contains the information that we have about the parameters based on the accumulated information from the past. The likelihood function specifies how plausible the observed data are given model parameter values. Therefore, defining a proper likelihood function is the central question in parameter estimation.

The parameter estimation problem is often written in a (cost) functional form with parameters as arguments.

Different numerical methods can be used for optimising the cost function and point estimates, such as the Maximum a Posteriori (MAP) estimate, can be obtained. In addition, Monte Carlo methods like MCMC can be used for producing samples from the posterior distribution instead of a single point estimate, and studying the uncertainty in the parameter values. For large-scale applications, applying MCMC is challenging, since the methods can involve thousands of repeated likelihood evaluations, but ways to improve the efficiency of MCMC for high-CPU models have been recently developed (Solonen et al., 2012).

## 2.3. Likelihood via filtering methods

In this section, the filter likelihood concept is discussed in the context of the extended Kalman filter and the ensemble adjustment Kalman filter. The approach is known in the parameter estimation of stochastic models (e.g. Singer, 2002), but less studied in connection with deterministic, chaotic systems (Hakkarainen et al., 2012). For the sake of completeness, the derivation of the likelihood computation is briefly reviewed in the Appendix.

When EKF is considered, the filter likelihood formula, i.e., the likelihood for observing  $y_{1:n}$  given the parameters  $\theta$ , can be written as

$$\begin{aligned} p(y_{1:n}|\theta) &= p(y_1|\theta) \prod_{k=2}^n p(y_k|y_{1:k-1}, \theta) \\ &= \prod_{k=1}^n \exp\left(-\frac{1}{2} r_k^T (C_k^y)^{-1} r_k\right) (2\pi)^{-d/2} |C_k^y|^{-1/2} \\ &\propto \exp\left(-\frac{1}{2} \sum_{k=1}^n \underbrace{[r_k^T (C_k^y)^{-1} r_k + \log |C_k^y|]}_{=J_k}\right), \end{aligned} \quad (3)$$

where  $r_k = y_k - \mathcal{H}(x_k^p)$  is the innovation vector and  $C_k^y = H_k C_k^p H_k^T + R_k$  is its error covariance matrix at time  $k$ . Operator  $|\cdot|$  denotes the matrix determinant. In filtering,  $\mathcal{H}$  is the observation operator that maps from the state space to the observation space and  $H$  is its linearization. The prior state  $x_k^p$  is given by the model, the prior covariance matrix  $C_k^p$  is given by the Kalman filtering formulas [eq. (1)] and  $R_k$  is the observation error covariance matrix.

Essentially, the filter likelihood formula [eq. (3)] depends on the summation over the cost functions  $J_k$ , where the normalising term  $\log |C_k^y|$  has to be accounted for, because it implicitly depends on the parameters  $\theta$ . The dependence comes from eq. (1) via the tangent-linear forecast model. It should be noted that eq. (3) itself does not require that  $x_k^p$  and its covariance matrix  $C_k^p$  are calculated using EKF. In principle, any state estimation method that produces these prior estimates could be used.

In EAKF, as implemented in the DART system, the discrepancy between the prior ensemble and the observations are calculated in the observation space for every single observation sequentially. The update of the ensemble and the weights are also calculated sequentially in the observation space. Thus, it is natural to adopt this approach also in the filter likelihood calculations. The cost function part of the filter likelihood formula [eq. (3)] can be calculated sequentially as a sum

$$J_k = \sum_{l=1}^L \frac{(y_l - \bar{z}_l^p)^2}{(\sigma_{y_l}^2 + \sigma_{z_l^p}^2)} + \log(\sigma_{y_l}^2 + \sigma_{z_l^p}^2), \quad (4)$$

where  $y_l$  is an individual observation and  $\bar{z}_l^p$  is the individual prior ensemble mean in observation space and  $\sigma_{y_l}^2$  and  $\sigma_{z_l^p}^2$  are their variances, respectively. The prior ensemble mean  $\bar{z}_l^p$  depends implicitly on  $y_{l-1}$ , because of the sequential update. The summation goes through all the individual observations  $y_{1:L,k}$  at time  $k$ . The DART system does not provide the above computation, but including these in the DART code can be easily implemented.

#### 2.4. Model and data

To explore the filter likelihood calculation, a modified version of the Lorenz-95 system is used (Lorenz, 1995; Wilks, 2005). The full system consists of two interrelated variables  $X$  and  $Y$ , and it is written as

$$\frac{dX_i}{dt} = -X_{i-1}(X_{i-2} - X_{i+1}) - X_i + F - \frac{hc}{b} \sum_{j=J(i-1)+1}^{Ji} Y_j \quad (5)$$

$$\frac{dY_j}{dt} = -cbY_{j+1}(Y_{j+2} - Y_{j-1}) - cY_j + \frac{c}{b}F_Y + \frac{hc}{b}X_{1+\lfloor \frac{j-1}{J} \rfloor} \quad (6)$$

where  $i = 1, \dots, I$  and  $j = 1, \dots, JI$ . That is, each of the ‘slow’ state variables  $X_i$  are forced by a sum of the additional fast variables  $Y_j$ . The fast variables have dynamics similar to the slow variables, but they are also coupled with the slow variables. In the model, cyclic boundary conditions, i.e.,  $X_{I+1} = X_1$  and  $Y_{JI+1} = Y_1$ , are used. In this paper, values  $I=40$ ,  $J=8$ ,  $F=F_Y=10$ ,  $h=1$  and  $c=b=10$ , adopted from Leutbecher (2010), are used.

The parametrized forecast model for the full system reads

$$\frac{dX_i}{dt} = -X_{i-1}(X_{i-2} - X_{i+1}) - X_i + F - g(X_i, \theta), \quad (7)$$

where  $g(X_i, \theta) = \theta_0 + \theta_1 X_i$  is the parameterization in which the effect of the missing ‘sub-grid’ scale fast variables  $Y_j$  are modeled using only the single, ‘resolved’, local variable  $X_i$ .

The aim of the experiment is to tune the parameters  $\theta = (\theta_0, \theta_1)$  given a synthetic set of observations  $y_{1:m}$ ,

computed by the full set of equations in (5) and (6) with additive noise, using the filter likelihood technique. In the experiment, one filtering step (‘24 h’) is 8 model integration steps (‘3 h’). Here the number of assimilation steps  $n=100$  (‘days’), and hence during each run the model is integrated  $100 \times 8$  times. The experiment is adopted from Hakkarainen et al. (2012) and a more comprehensive introduction is given there.

#### 2.5. Experiment setup

In the experiment, the parameter values are sampled from the posterior distribution {consisting of filter likelihood [eq. (3)] and uninformative prior} using an adaptive Metropolis algorithm (Haario et al., 2001). The chain length is selected to be 3000 and the first 500 samples (the burn-in period) are discarded. The filter likelihood MCMC experimentation is synthesised with the following pseudo-algorithm, where the superscript  $m$  on  $\theta$  denotes MCMC chain index:

Step 0:

Initialise the MCMC run and select  $\theta^1$ .

Step 1:

Propose a new candidate  $\hat{\theta}$  for the parameters.

Step 2:

Run the state estimation system using  $\hat{\theta}$  and evaluate formula [eq. (3)] during the run.

Step 3:

Accept  $\hat{\theta}$  with probability  $\min\left(1, \frac{p(y_{1:n}|\hat{\theta})p(\theta)}{p(y_{1:n}|\theta^m)p(\theta^m)}\right)$ .

Step 4:

If  $\hat{\theta}$  is accepted set  $\theta^{m+1} = \hat{\theta}$ , else set  $\theta^{m+1} = \theta^m$ .

Step 5:

Set  $m=m+1$  and go to step 1 until  $m$  is equal to the chain length.

In our implementation, we use an uninformative (flat) prior, which means that the prior is a constant,  $p(\theta) \propto 1$ . Then, the prior term cancels out in Step 3 of the above algorithm, since  $p(\hat{\theta})/p(\theta^m) = 1$ .

For readers not familiar with MCMC, a short introduction is in order. The MCMC algorithm works by generating candidate values from a *proposal distribution*, which in our case is Gaussian centred at the current parameter value. The candidates are then either accepted or rejected according to a simple rule that contains the ratio of the posterior densities at the proposed point and at the current point, see Step 3 of the above algorithm. One can see that moves ‘upward’ (to a point with higher posterior density) are always accepted. However, also moves downward can be accepted, and the resulting random walk ends up exploring the posterior distribution without converging to a single point. The algorithm is a standard tool in statistical

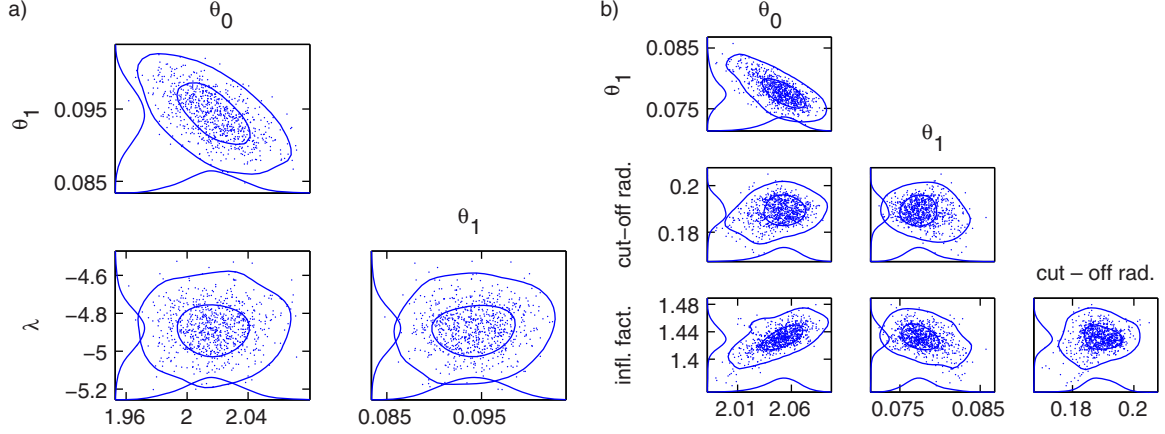


Fig. 1. Pairwise MCMC samples using EKF (left panel) and EAKF (right panel) version of the filter likelihood formula [eq. (3)]. In addition to the model parameters  $\theta_0$  and  $\theta_1$ , filter related parameters are sampled too. Contours illustrate 95% and 50% probability regions. Also, the one-dimensional marginal densities are illustrated.

analysis and can be shown to yield samples from the correct target density. Here, we apply an adaptive version of the algorithm Haario et al. (2001), where the proposal distribution is automatically tuned to better match the posterior distribution. For more details on MCMC, see Robert and Casella (2005).

Using the filter likelihood [eq. (3)] for MCMC, we are able to estimate both the model parameters  $\theta$  and the filter-specific tuning parameters. In the experiments, for the EKF version, the model error covariance matrix is parameterised as follows

$$Q = \exp(\lambda)I, \quad (8)$$

where  $I$  is an identity matrix. The parameter  $\lambda$  is tuned together with the model parameters  $\theta_0$  and  $\theta_1$ . In the EAKF version of the filter likelihood, the prior covariance inflation factor and the cut-off radius related to localisation (Gaspari and Cohn, 1999), defined in the DART setup, are tuned together with the model parameters  $\theta$ .

As a validation metric for the parameters, we use the average root mean squared error (RMSE) defined as

$$r(\theta) = \frac{1}{n} \sum_{k=1}^n \sqrt{\frac{1}{I} \sum_{i=1}^I (x_{i,k}^{\text{est}}(\theta) - x_{i,k}^{\text{true}})^2} \quad (9)$$

where  $x_k^{\text{true}}$  is the true state and  $x_k^{\text{est}}$  is the posterior estimate. We evaluate RMSE on a fixed grid for  $\theta_1$  and  $\theta_2$  keeping the filter-specific parameters fixed.

## 2.6. Results

The experiments are started by a step, where all parameters (i.e. filter parameters and the model parameters) are estimated together with MCMC. In Fig. 1, pairwise MCMC

samples using EKF (left panel) and EAKF (right panel) are shown. It can be seen, that the marginal distributions are rather Gaussian and all parameters are identified. Based on the mean values of the MCMC chains, values  $-4.8824$ ,  $1.4310$  and  $0.1889$  were chosen for  $\lambda$ , the prior inflation factor and the cut-off radius, respectively.

In the second step, to further underline the difference in the model parameter estimates in the usual situation where the filter parameters are fixed, we calculate the conditional  $\theta$  distributions using MCMC by fixing the filter parameters to the mean values found in the previous step. For the validation purposes, the RMS error [eq. (9)] is evaluated on a fixed grid to explore the parameter space for ‘all’ possible pairs of  $\theta_0$  and  $\theta_1$ .

The results are illustrated in Fig. 2, where the 50% and 95% probability contours calculated from the EKF and EAKF MCMC chains are plotted using solid and dashed lines, respectively. The colours on the background indicate the average RMSE values [eq. (9)] for EKF (left panel) and EAKF (right panel). We observe that the closure parameter values that correspond to the lowest average RMSE values for EKF and EAKF are not the same. This implies that the optimal parameter values for EKF are not optimal for EAKF and vice versa. This shows that there is a dilemma of the uniqueness of the model closure parameters: an identical model formulation can have more than one optimal parameter value combination. This can be explained by the fact that the two data assimilation components have different systematic approximation errors, and these errors are compensated, to some extent, by the different parameter values in the forecast model. In this sense, the forecast model parameter values should be considered as filter-dependent quantities. The consequences of this result are discussed below in Section 3.

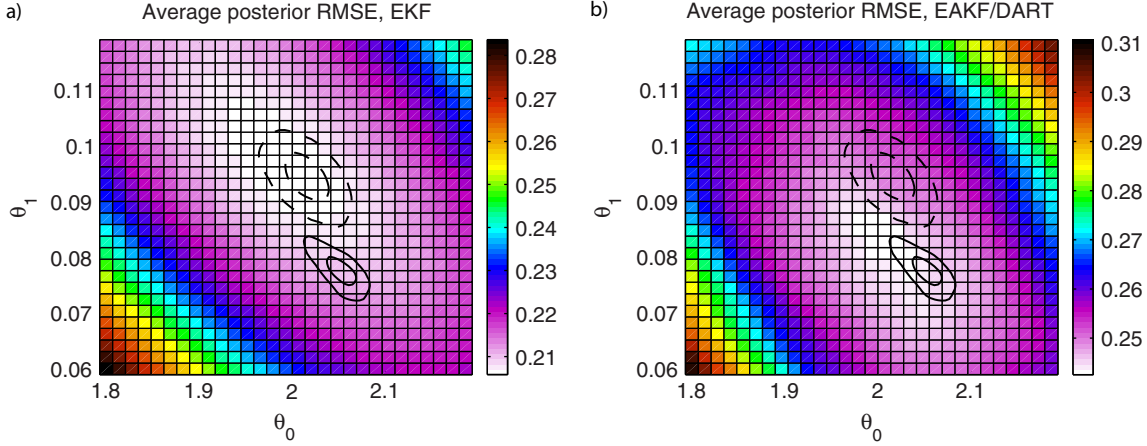


Fig. 2. Average posterior RMSE values [eq. (9)] when using EKF (left panel) and EAKF (right panel). Lower values indicate better performance. Solid and dashed lines indicate the 50% and 95% probability regions as obtained with MCMC in case of EAKF and EKF, respectively. RMSE values are depicted by the background colour scheme. Note the difference in colour scales.

Figure 2 also depicts that the filter likelihood method captures the optimal parameter values correctly: the lowest posterior RMSE values are close to the obtained posterior distributions. In addition, it can be seen that the results produced by the EKF and EAKF versions of the filter likelihood technique are similar, but not identical. The posterior distribution obtained using EKF is closer to the parameter values that yield low RMS error for EKF. And vice versa: the parameters estimated using EAKF are closer to the optimal values for EAKF than the ones obtained by EKF. That is, applying the filter likelihood technique seems to provide rather good filter-specific parameter estimates. In this study, the RMSE values produced by EKF are lower than those produced by EAKF (see the RMSE values at the background of Fig. 2), which was expected. In addition, it can be noted, that the choice of data assimilation component has a much larger effect on the RMSE than the exact values of the model parameters.

### 3. Discussion and conclusions

In this article, we presented a numerical experiment with the purpose to simulate the prediction model parameter estimation problem in chaotic systems. We studied a simplified prediction system, which consists of a modified Lorenz-95 system as the prediction model, where the net effect of fast variables is represented by a two-parameter linear scheme. Two versions of the Kalman filter were used as different data assimilation components. We have shown that depending on the version of data assimilation that is applied, either the extended Kalman filter or the ensemble adjustment Kalman filter, the optimal prediction model

tuning parameters corresponding to the highest system accuracy have two distinct optima. Accuracy is measured here as the RMS errors of the state estimate against the known truth. Replacement of EKF with EAKF implies a major change to the prediction system and it is only natural that the prediction model parameters need re-tuning.

In addition, we have demonstrated a method that can be used to estimate the closure parameters for each prediction system component separately. The approach is based on computing the likelihood using the output of the data assimilation system. We also demonstrate how Markov chain Monte Carlo sampling methods can be used on top of this likelihood formulation. The method recovers the optimal forecast model parameter values for both data assimilation systems. In addition, both the forecast model parameters and the data assimilation related parameters can be estimated. In our case, the data assimilation related parameters were the model error covariance matrix in EKF, and covariance inflation and localisation parameters in EAKF.

The models applied in operational weather forecasting are far more complex than the system applied here. If we assume that the behaviour discovered here scales up, it has some consequences for tuning of large-scale systems. Traditionally, model releases are carefully fine-tuned and parallel tested associated with forecast model changes. Changes in data assimilation are, however, typically considered independent of forecast model fine-tuning. According to the result presented here, this should not be the case. If the data assimilation component undergoes a major revision but the prediction model remains the same, then the prediction model tuning corresponds to the previous

version of the data assimilation component and may thus be suboptimal. In this case, the correct procedure would be to consider model tuning, too.

Finally, this work considered a low-order prediction system. Our future work will be directed towards quantifying how this result scales-up to more realistic systems. We will use the ECHAM6 prediction model together with the DART toolbox to study this question.

#### 4. Acknowledgements

We would like to acknowledge the National Center for Atmospheric Research for making the Data Assimilation Research Testbed available on-line. The research has been financially supported by the Academy of Finland (project numbers 127210, 132808, 133142, 134935 and 134999) and the Centre of Excellence in Inverse Problems Research. Additionally, this work was funded by the European Commission's 7th Framework Programme, under Grant Agreement number 282672, EMBRACE project, and the Nessling Foundation.

#### 5. Appendix

##### A.1. Derivation of the filter likelihood formula

We present the derivation of the filter likelihood formula [eq. (3)], briefly. A more comprehensive discussion is given in Hakkarainen et al. (2012).

Let us consider the following general state space model.

$$x_k \sim p(x_k | x_{k-1}, \theta) \quad (\text{A1})$$

$$y_k \sim p(y_k | x_k) \quad (\text{A2})$$

$$\theta \sim p(\theta) \quad (\text{A3})$$

of state  $x_k$ , observations  $y_k$  and parameters  $\theta$ , respectively. Our goal is to find the posterior distribution  $p(\theta | y_{1:n})$ . In filtering, the previous state estimate can be transported to the next time step's prior by using the prediction model

$$p(x_k | y_{1:k-1}, \theta) = \int p(x_k | x_{k-1}, \theta) \times p(x_{k-1} | y_{1:k-1}, \theta) dx_{k-1} \quad (\text{A4})$$

and in the analysis step the prior is updated by  $y_k$  using the Bayes' formula

$$p(x_k | y_{1:k}, \theta) \propto p(y_k | x_k, \theta) p(x_k | y_{1:k-1}, \theta). \quad (\text{A5})$$

The predictive distribution for next observations can be obtained by

$$p(y_k | y_{1:k-1}, \theta) = \int p(y_k | x_k, \theta) p(x_k | y_{1:k-1}, \theta) dx_k. \quad (\text{A6})$$

Using Bayes' formula and the chain rule we obtain

$$\begin{aligned} p(\theta | y_{1:n}) &\propto p(y_{1:n} | \theta) p(\theta) = p(y_n | y_{1:n-1}, \theta) \\ &\quad \times p(y_{n-1} | y_{1:n-2}, \theta) \cdots p(y_2 | y_1, \theta) \\ &\quad \times p(y_1 | \theta) p(\theta), \end{aligned} \quad (\text{A7})$$

where  $p(y_k | y_{1:k-1}, \theta)$  can be calculated using eq. (A6).

In extended Kalman filtering the predictive distribution is

$$y_k | y_{1:k-1}, \theta \sim N(\mathcal{H}(x_k^p), C_k^y). \quad (\text{A8})$$

Now, applying this to eq. (A8) we can obtain eq. (3).

#### References

- Anderson, J., Hoar, T., Raeder, K., Liu, H., Collins, N. and co-authors. 2009. The data assimilation research testbed: a community facility. *B. Am. Meteorol. Soc.* **90**(9), 1283–1296. DOI: 10.1175/2009BAMS2618.1.
- Anderson, J. L. 2001. An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.* **129**(12), 2884–2903. DOI: 10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO.
- Anderson, J. L. 2010. A non-Gaussian ensemble filter update for data assimilation. *Mon. Wea. Rev.* **138**(11), 4186–4198. DOI: 10.1175/2010MWR3253.1.
- Anderson, J. L. and Anderson, S. L. 1999. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.* **127**(12), 2741–2758. DOI: 10.1175/1520-0493(1999)127<2741:AMCIOT>2.0.CO.
- Annan, J. D., Lunt, D. J., Hargreaves, J. C. and Valdes, P. J. 2005. Parameter estimation in an atmospheric GCM using the Ensemble Kalman Filter. *Nonlin. Processes Geophys.* **12**(3), 363–371. DOI: 10.5194/npg-12-363-2005.
- D'Andrea, F. and Vautard, R. 2000. Reducing systematic errors by empirically correcting model errors. *Tellus A* **52**(1), 21–41. DOI: 10.1034/j.1600-0870.2000.520103.x.
- Dowd, M. 2011. Estimating parameters for a stochastic dynamic marine ecological system. *Environmetrics* **22**(4), 501–515. DOI: 10.1002/env.1083.
- Gaspari, G. and Cohn, S. E. 1999. Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteorol. Soc.* **125**(554), 723–757. DOI: 10.1002/qj.49712555417.
- Gelman, A., Carlin, J., Stern, H. and Rubin, D. 2003. *Bayesian Data Analysis*. 2nd ed. Chapman & Hall, Boca Raton, FL.
- Haario, H., Saksman, E. and Tamminen, J. 2001. An adaptive Metropolis algorithm. *Bernoulli* **7**(2), 223–242.
- Hakkarainen, J., Ilin, A., Solonen, A., Laine, M., Haario, H. and co-authors. 2012. On closure parameter estimation in chaotic systems. *Nonlin. Processes Geophys.* **19**(1), 127–143. DOI: 10.5194/npg-19-127-2012.
- Hamill, T. M., Whitaker, J. S. and Snyder, C. 2001. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.* **129**(11), 2776–2790. DOI: 10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO.

- Jakob, C. 2010. Accelerating progress in global atmospheric model development through improved parameterizations: challenges, opportunities, and strategies. *B. Am. Meteorol. Soc.* **91**(7), 869–875. DOI: 10.1175/2009BAMS2898.1.
- Järvinen, H., Räisänen, P., Laine, M., Tamminen, J., Ilin, A. and co-authors. 2010. Estimation of ECHAM5 climate model closure parameters with adaptive MCMC. *Atmos. Chem. Phys.* **10**(20), 9993–10002. DOI: 10.5194/acp-10-9993-2010.
- Jung, T. 2005. Systematic errors of the atmospheric circulation in the ECMWF forecasting system. *Quart. J. Roy. Meteorol. Soc.* **131**(607), 1045–1073. DOI: 10.1256/qj.04.93.
- Kaas, E., Guldberg, A., May, W. and Déqué, M. 1999. Using tendency errors to tune the parameterisation of unresolved dynamical scale interactions in atmospheric general circulation models. *Tellus A* **51**(5), 612–629. DOI: 10.1034/j.1600-0870.1999.00006.x.
- Kalman, R. E. 1960. A new approach to linear filtering and prediction problems. *Trans. ASME J. Basic Eng. Ser. D.* **82**, 35–42.
- Klinker, E. and Sardeshmukh, P. D. 1992. The diagnosis of mechanical dissipation in the atmosphere from large-scale balance requirements. *J. Atmos. Sci.* **49**(7), 608–627. DOI: 10.1175/1520-0469(1992)049<0608:TDOMDI>2.0.CO.
- Kunz, T., Fraedrich, K. and Kirk, E. 2008. Optimisation of simplified GCMs using circulation indices and maximum entropy production. *Clim. Dynam.* **30**, 803–813. DOI: 10.1007/s00382-007-0325-y.
- Leutbecher, M. 2010. Predictability and ensemble forecasting with Lorenz-95 systems, Lecture notes. ECMWF meteorological training course on Predictability, diagnostics and seasonal forecasting. Online at: [http://www.ecmwf.int/newsevents/training/meteorological\\_presentations/pdf/PR/Practice\\_L95.pdf](http://www.ecmwf.int/newsevents/training/meteorological_presentations/pdf/PR/Practice_L95.pdf)
- Lohmann, U., Quaas, J., Kinne, S. and Feichter, J. 2007. Different approaches for constraining global climate models of the anthropogenic indirect aerosol effect. *B. Am. Meteorol. Soc.* **88**(2), 243–249. DOI: 10.1175/BAMS-88-2-243.
- Lorenz, E. N. 1995. Predictability: a problem partly solved. In: *Proceedings of the Seminar on Predictability, European Center on Medium Range Weather Forecasting* **1**, 1–18. Online at: <http://www.ecmwf.int/publications/library/do/references/show?id=87423>
- Neelin, J. D., Bracco, A., Luo, H., McWilliams, J. C. and Meyerson, J. E. 2010. Considerations for parameter optimization and sensitivity in climate models. *Proc. Natl. Acad. Sci.* **107**(50), 21349–21354. DOI: 10.1073/pnas.1015473107.
- Robert, C. P. and Casella, G. 2005. *Monte Carlo Statistical Methods (Springer Texts in Statistics)*. Springer-Verlag, Secaucus, NJ.
- Schwartz, C. S., Liu, Z., Chen, Y. and Huang, X.-Y. 2011. Impact of assimilating microwave radiances with a limited-area ensemble data assimilation system on forecasts of typhoon morakot. *Wea. Forecasting* **27**(2), 424–437. DOI: 10.1175/WAF-D-11-00033.1.
- Severijns, C. A. and Hazeleger, W. 2005. Optimizing parameters in an atmospheric general circulation model. *J. Climate* **18**(17), 3527–3535. DOI: 10.1175/JCLI3430.1.
- Singer, H. 2002. 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.
- Solonen, A., Ollinaho, P., Laine, M., Haario, H., Tamminen, J. and co-authors. 2012. Efficient MCMC for climate model parameter estimation: parallel adaptive chains and early rejection. *Bayesian Anal.* **7**(3), 715–736. DOI: 10.1214/12-BA724.
- Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M. and Whitaker, J. S. 2003. Ensemble square root filters. *Mon. Wea. Rev.* **131**(7), 1485–1490. DOI: 10.1175/1520-0493(2003)131<1485:ESRF>2.0.CO.
- Torn, R. D. and Davis, C. A. 2012. The influence of shallow convection on tropical cyclone track forecasts. *Mon. Wea. Rev.* **140**(7), 2188–2197. DOI: 10.1175/MWR-D-11-00246.1.
- Wilks, D. 2005. Effects of stochastic parametrizations in the Lorenz 96 system. *Quart. J. Roy. Meteorol. Soc.* **131**(606), 389–407. DOI: 10.1256/qj.04.03.