Department of Mathematics and Statistics

Preprint MPS-2013-14

7 November 2013

2 Estimating the observation error covariance matrix with the ensemble transform Kalman filter

Data assimilation techniques combine observations $\mathbf{y}_n \in \mathbb{R}^{N^p}$ at time t_n with a model prediction of the state, the background $\mathbf{x}_n^f \in \mathbb{R}^{N^m}$, which is often determined by a previous forecast. The observations and background are weighted by their respective errors, to provide a best estimate of the state $\mathbf{x}_n^a \in \mathbb{R}^{N^m}$, known as the analysis. This analysis is then forecast using the possibly non-linear model M_n to provide a background at the next assimilation time,

$$\mathbf{x}_{n+1}^f = M_n(\mathbf{x}_n^a). \tag{1}$$

We now give a brief overview of the ensemble transform Kalman filter (ETKF) [Bishop et al., 2001, Livings et al., 2008] that we will adapt and the notation that is used in this study. At time t_n we have an ensemble, a statistical sample of N state estimates $\{\mathbf{x}_n^i\}$ for $i = 1 \dots N$. These ensemble members are stored in a state ensemble matrix $\mathbf{X}_n \in \mathbb{R}^{N^m \times N}$ where each column of the matrix is a state estimate for an individual ensemble member,

$$\mathbf{X}_n = \begin{pmatrix} \mathbf{x}_n^1 & \mathbf{x}_n^2 & \dots & \mathbf{x}_n^N \end{pmatrix}. \tag{2}$$

It is possible to calculate the ensemble mean,

$$\bar{\mathbf{x}}_n = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_n^i, \tag{3}$$

and subtracting the ensemble mean from the state ensembles gives the ensemble perturbation matrix

$$\mathbf{X}_n' = \begin{pmatrix} \mathbf{x}_n^1 - \bar{\mathbf{x}}_n & \mathbf{x}_n^2 - \bar{\mathbf{x}}_n & \dots & \mathbf{x}_n^N - \bar{\mathbf{x}}_n \end{pmatrix}. \tag{4}$$

This allows us to write the ensemble covariance matrix as

$$\mathbf{P}_n = \frac{1}{N-1}\mathbf{X}_n'\mathbf{X}_n'^T. \tag{5}$$

For the ensemble transform Kalman filter (ETKF) the analysis at time t_n is given by,

$$\bar{\mathbf{x}}_n^a = \bar{\mathbf{x}}_n^f + \mathbf{K}_n(\mathbf{y}_n - H_n(\bar{\mathbf{x}}_n^f)), \tag{6}$$

where $\bar{\mathbf{x}}_n^a$ is the analysis ensemble mean and $\bar{\mathbf{x}}_n^f$ is the forecast ensemble mean. The possibly non-linear observation operator $H : \mathbb{R}^{N^p} \to \mathbb{R}^{N^m}$ maps the state space to the observation space . The Kalman gain matrix,

$$\mathbf{K}_n = \mathbf{P}_n^f\mathbf{H}_n^T(\mathbf{H}_n\mathbf{P}_n^f\mathbf{H}_n^T + \mathbf{R}_n)^{-1}, \tag{7}$$

is a matrix of size $N^m \times N^p$ where $\mathbf{H}_n$ is the observation operator linearised about the background state. The observation error covariance matrix is denoted by $\mathbf{R}_n \in \mathbb{R}^{N^p \times N^p}$ and $\mathbf{P}_n^f \in \mathbb{R}^{N^m \times N^m}$ is the forecast error covariance matrix. When the forecast error covariance is derived from climatological data and assumed static, it is often denoted as $\mathbf{B}_n$ and known as the background error covariance matrix.

Previously it has been assumed that the observation error covariance matrix $\mathbf{R}$ is diagonal. However, with recent work showing that $\mathbf{R}$

nb K

2.1 The DBCP diagnostic

3.2.1 The observations

To create observations we must add errors from a specified dis

KS equation are given in Cox and Matthews [2000] and Kassam and Trefethen [2005]. The truth is defined by the solution to the KS equation on the periodic domain 0 spread. If the ensemble spread is not maintained the analysis and the estimation of the observation error covariance matrix may be affected.

4.1 Results with a static R and frequent observations

We begin by considering the case when the true observation error covariance matrix is static.

In Experiments 1L and 1K we use the standard EKTF for the assimilation. We begin by setting the true matrix R^t to $R^t = R^D + R^C$, where $R^D = 0.1I$ and $R^C = 0.1C$. The Lorenz '96 and KS models are each run for 1000 assimilation steps. We assume that R is diagonal, with $R_0 = \mathrm{diag}(R^t)$. The standard ETKF is used to gain an estimate of R after the assimilation. The background and analysis innovations are calculated throughout the assimilation window. After the final assimilation these

covariance matrix to be calculated.

We showed it is possible to obtain a good estimate of R using the DBCP diagnostic. We then showed that estimating R within the ETKF worked well, with good estimates obtained, the ensemble spread maintained and the analysis RMSE reduced compared to the case where the matrix R is always assumed diagonal. We also showed that the method does not work as well where the observations are less frequent, although this may be dependent on the model. However the method still produces a reasonable estimate of R, maintains the ensemble variance and the time-averaged analysis RMSE is lower than where a diagonal R is used.

We next considered a case where R varied slowly with time. We showed that the method worked well where the true R was defined to slowly vary with time. The time-averaged analysis RMSE was low and the ensemble spread was maintained. The estimates of the correlation structure were good, suggesting that the method is capable of estimating a slowly time-varying observation error covariance matrix. A case where the length-scale of the observation error covariance varied more quickly was also considered, and the ETKFR produced reasonable estimates of the observation error covariance matrix. We also showed that the ability of the method to approximate the correlation structure was not sensitive to the forecast error variances or the true magnitude of the observation error variance. This suggests that the method would be suitable to give a time-dependent estimate of correlated observation error. We note that the effectiveness of the method will depend on how rapidly the synoptic situation and hence correlated error is changing and how often observations are available. The correlated error will also be dependent on the dynamical system. For models designed to capture rapidly developing situations, where representativity error and hence correlated error is likely to change rapidly, assimilation cycling and observation frequency within the assimilation is expected to be more frequent and hence more data is available for estimating the observation error.

F. Hilton, A. Collard, V. Guidard, R. Randriamampianina, an d M. Schwaerz. Assimilation of IASI radiances at european NWP centres. In Proceedings of Workshop on the assimilation of IASI data in NWP, ECMWF, Reading, UK, 6-8 May 20 09, pages 39{48, 2009.

A. Hollingsworth and P. Lönnberg. The statistical structu re of short-range forecast errors as determined from radiosonde data. Part I: The wind eld. Tellus, 38A:111{136, 1986.

P. L. Houtekamer and H. L. Mitchell. Ensemble Kalman lterin g. Quarterly Journal of the Royal Meteorological Society 133:3260{3289, 2005.

T. Janjic and S. E. Cohn. Treatment of observation error due to unresolved scales in atmospheric data assimilation. Monthly Weather Review, 134:2900{2915, 2006.

M. Jardak, I. M. Navon, and M. Zupanski. Comparison of sequential data assimilation methods for the Kuramoto-Sivashinsky equation. International Journal for Numerical Methods in Fluids, 62:374{402, 2000.

A. Kassam and L. Trefethen. Fourth-order time-stepping for sti pdes. SIAM J. Sci. Computing, 25:1214{1233, 2005.

Y. Kuramoto. Di usion-induced chaos in reaction systems. Progress of Theoretical Physics, 64:346{367, 1978.

L. M. Stewart, S. L. Dance, and N. K. Nichols. Correlated observation errors in data assimilation. International Journal for Numerical Methods in Fluids, 56:1521–1527, 2008.

L. M. Stewart, J. Cameron, S. L. Dance, S. English, J. R. Eyre, and N. K. Nichols. Observation error correlations in IASI radiance data. Technical report, University of Reading, 2009. Mathematics reports series, www.reading.ac.uk/web/FILES/maths/obs_error_IASI_radiance.pdf.

L. M. Stewart, S. L. Dance, and N. K. Nichols. Data assimilation with correlated observation errors: experiments with a 1-D shallow water model. Tellus A, 65, 2013a.

L. M. Stewart, S. L. Dance, N. K. Nichols, J. R. Eyre, and J. Cameron. Estimating interchannel observation-error correlations for IASI radiance data in the Met Office system. Quarterly Journal of the Royal Meteorological Society, 2013b. doi: 10.1002/qj.2211.

H. Thiebaux. Anisotropic correlation functions for objective analysis. Monthly Weather Review, 104:994–1002, 1976.

J. A. Waller. Using observations at different spatial scales in data assimilation for environmental prediction. PhD thesis, University of Reading, Department of Mathematics and Statistics, 2013. http://www.reading.ac.uk/maths-and-stats/research/theses/maths-phdtheses.aspx.

J. A. Waller, S. L. Dance, A. S. Lawless, N. K. Nichols, and J. R Eyre. Representativity error for temperature and humidity using the Met Office UKV model. Quarterly Journal of the Royal Meteorological Society, 2013. Early View. DOI: 10.1002/qj.2207.

P. Weston. Progress towards the implementation of correlated observation errors in 4d-var. Technical report, Met Office, UK, 2011. Forecasting Research Technical Report 560.

J. S. Whitaker, T. M. Hamill, X. Wei, Y. Song, and Z. Toth. Ensemble data assimilation with the NCEP global forecast system. Monthly Weather Review, 136:463–482, 2008.

List of Figures

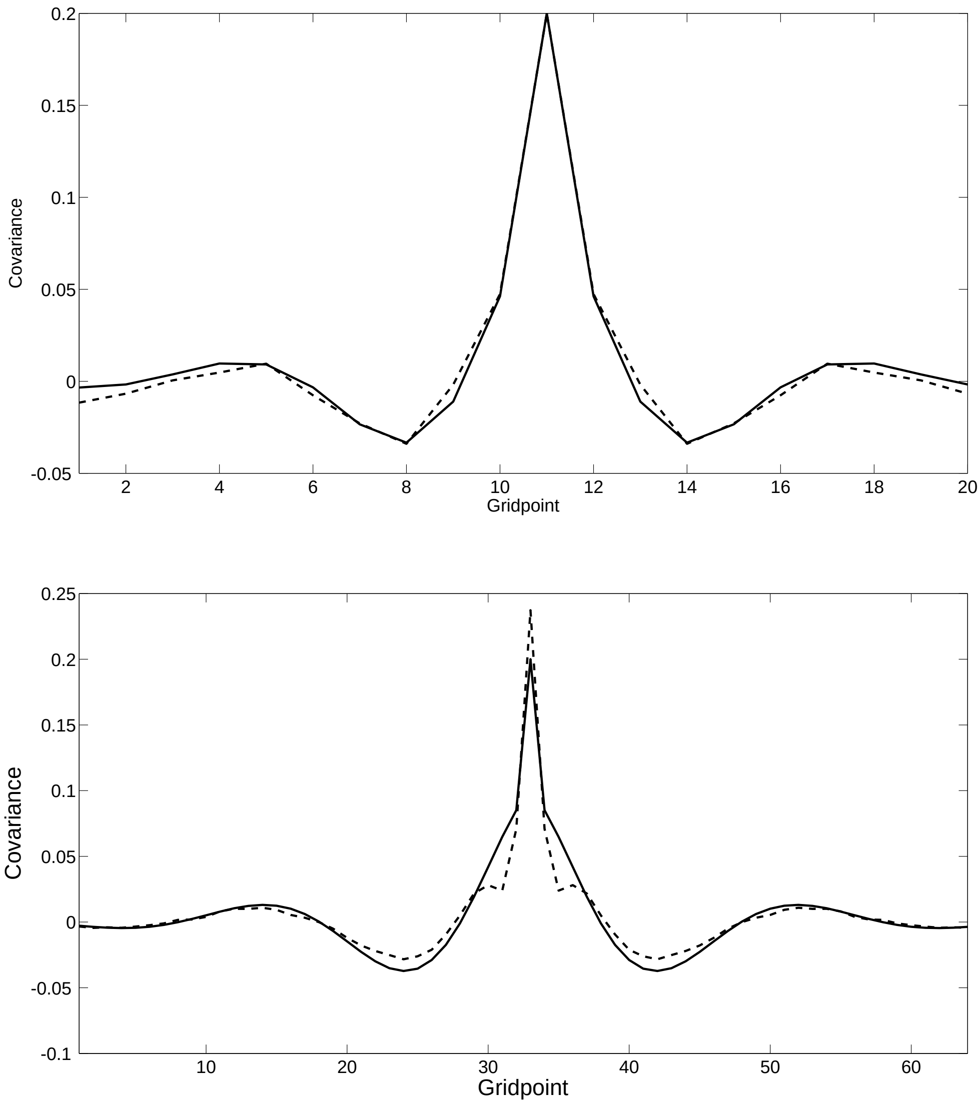
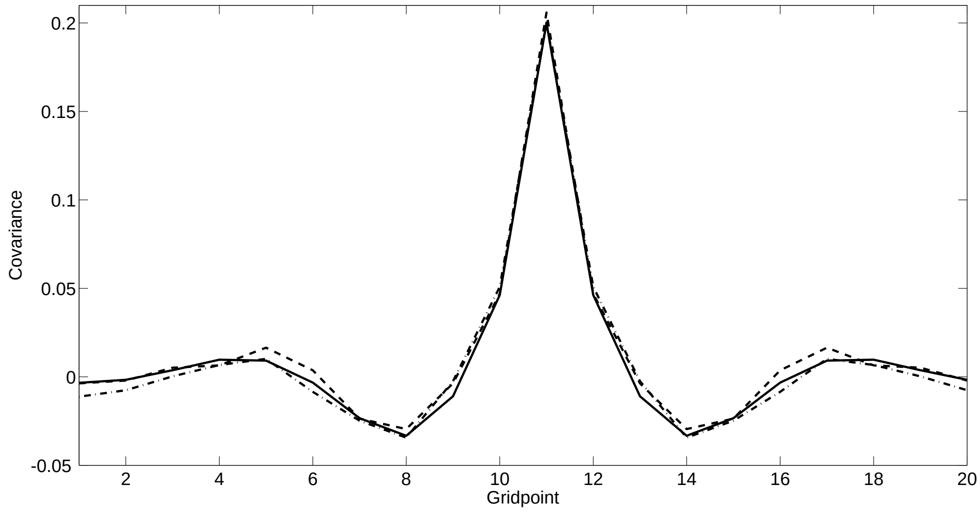


Figure 1 { Rows of the true (solid) and estimated (dashed) covariance matrices a) Experiment 1L. Observation error covariance RMSE: 0.002. b) Experiment 1K. Observation error covariance RMSE: 0.010.


Covariance
0.2
0.15
0.1
0.05
0
-0.05
2
4
6
8
10
12
14
16
18
20
Gridpoint

List of Tables

Table 1 { Details of experiments executed using the Lorenz '96 model to investigate the performance of the ETKF with observation error covariance estimation

Exp. No.	True R	Assimilation Method	Obs Freq (time steps)	$^{2}_{b}$, $^{2}_{D}$, $^{2}_{C}$	Time Av analysis	Covariance RMSE

Table 2