Department of Meteorology

University of Reading

Incorporating Correlated Observation Errors in Ensemble Data Assimilation

J. A. Waller | S. L. Dance | A. S. Lawless | N. K. Nichols

Introduction

Data assimilation techniques combine observations with a model prediction of the state, the background, to provide a best estimate of the state known as the analysis. The errors associated with the observations can be attributed to a number of sources, including representativity error which is correlated and state dependent [1].

Until recently the observation error covariance matrix has been assumed uncorrelated. However, it has been shown that the inclusion of the correlated errors in the assimilation leads to: a more accurate analysis, the inclusion of more observation information content and an improvement in the NWP skill score [2].

Aims

We aim to:

- Estimate a time-varying observation error covariance matrix, $\boldsymbol{R}$, within an ensemble data assimilation system.
- Improve the estimated states of the system given by the analysis.

The diagnostic of Desroziers et al. [2005]

Desroziers et. al. [3] show that $\boldsymbol{R}$ can be estimated using the diagnostic,

$$\boldsymbol{R} \approx E[(\boldsymbol{y} - \boldsymbol{H}(\boldsymbol{x}^b))(\boldsymbol{y} - \boldsymbol{H}(\boldsymbol{x}^a))^T] \quad (1)$$

The diagnostic is valid if $\boldsymbol{B}$ and $\boldsymbol{R}$ used in the assimilation are exact. It has been used successfully to estimate $\boldsymbol{R}$ and produced good results even when $\boldsymbol{R}$ and $\boldsymbol{B}$ used in the assimilation are not exact.

The ETKF with R estimation (ETKFR)

We introduce the following new algorithm for estimating $\boldsymbol{R}$.

Initialisation Determine the initial ensemble and error covariance matrix $\boldsymbol{R}_0$.

Step 1 Forecast each ensemble member, determine the ensemble mean, $\bar{\boldsymbol{x}}^f$, and perturbation matrix.

Step 2 Calculate the background innovations, $\boldsymbol{d}^b = \boldsymbol{y}_k - \boldsymbol{H}(\bar{\boldsymbol{x}}_k^f)$

Step 3 Update the ensemble mean and perturbations.

Step 4 Calculate the analysis innovations, $\boldsymbol{d}^a = \boldsymbol{y}_k - \boldsymbol{H}(\bar{\boldsymbol{x}}_k^a)$

Step 5 If $k > N^s$ update R using

$$\boldsymbol{R}_{k+1} \approx \frac{1}{N^s-1}\sum_{j=N^s-k}^{j=k}\left(\boldsymbol{y}_j - \boldsymbol{H}(\bar{\boldsymbol{x}}_j^a)\right)\left(\boldsymbol{y}_j - \boldsymbol{H}(\bar{\boldsymbol{x}}_j^b)\right)^T$$

Step 6 Symmetrise R,

$$\boldsymbol{R}_{k+1} = \frac{1}{2}\left(\boldsymbol{R}_{k+1} + \boldsymbol{R}_{k+1}^T\right)$$

Twin Experiments

To analyse the ensemble transform Kalman filter with $\boldsymbol{R}$ estimation (ETKFR) we run a series of twin experiments. From a truth run we create direct pseudo-observations with observation error covariance matrix $\boldsymbol{R} = \boldsymbol{R}^D + \boldsymbol{R}^C$, the sum of uncorrelated and correlated errors (Figure 1).

For the correlated errors we draw from the SOAR function. To create time dependent error correlations the length scale of the SOAR function is varied with time.

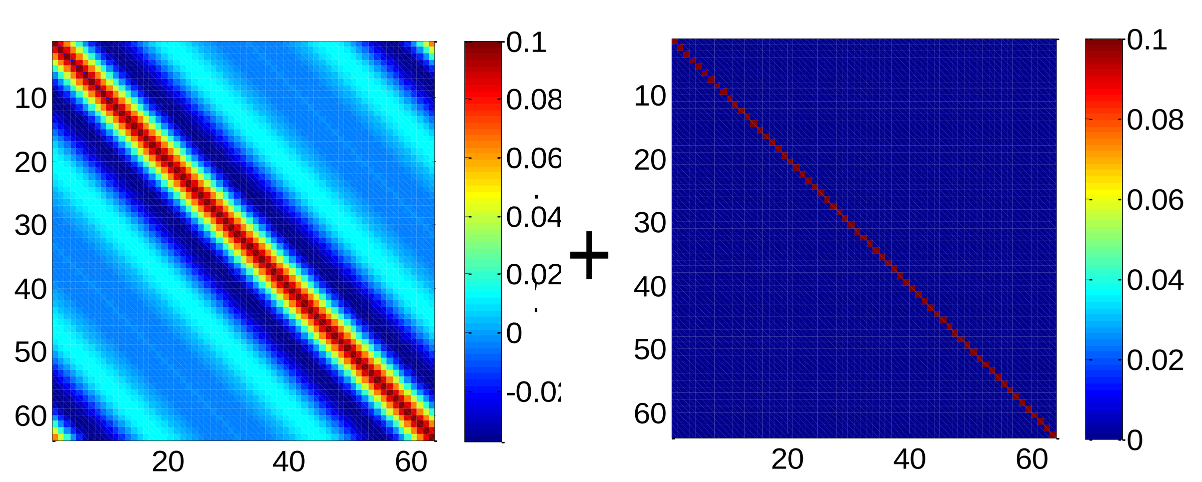

Figure 1: The correlated (left) and diagonal (right) contributions to the observation error covariance matrix with the correlated and diagonal error variances set to 0.1

Results

Experiments have been carried out using the Lorenz '96 and Kuramoto Sivashinsky models [4].

Using the method:

- The varying error covariance structure is captured (Figure 2).
- The analysis is improved by including the estimated $\boldsymbol{R}$ matrix in the assimilation.

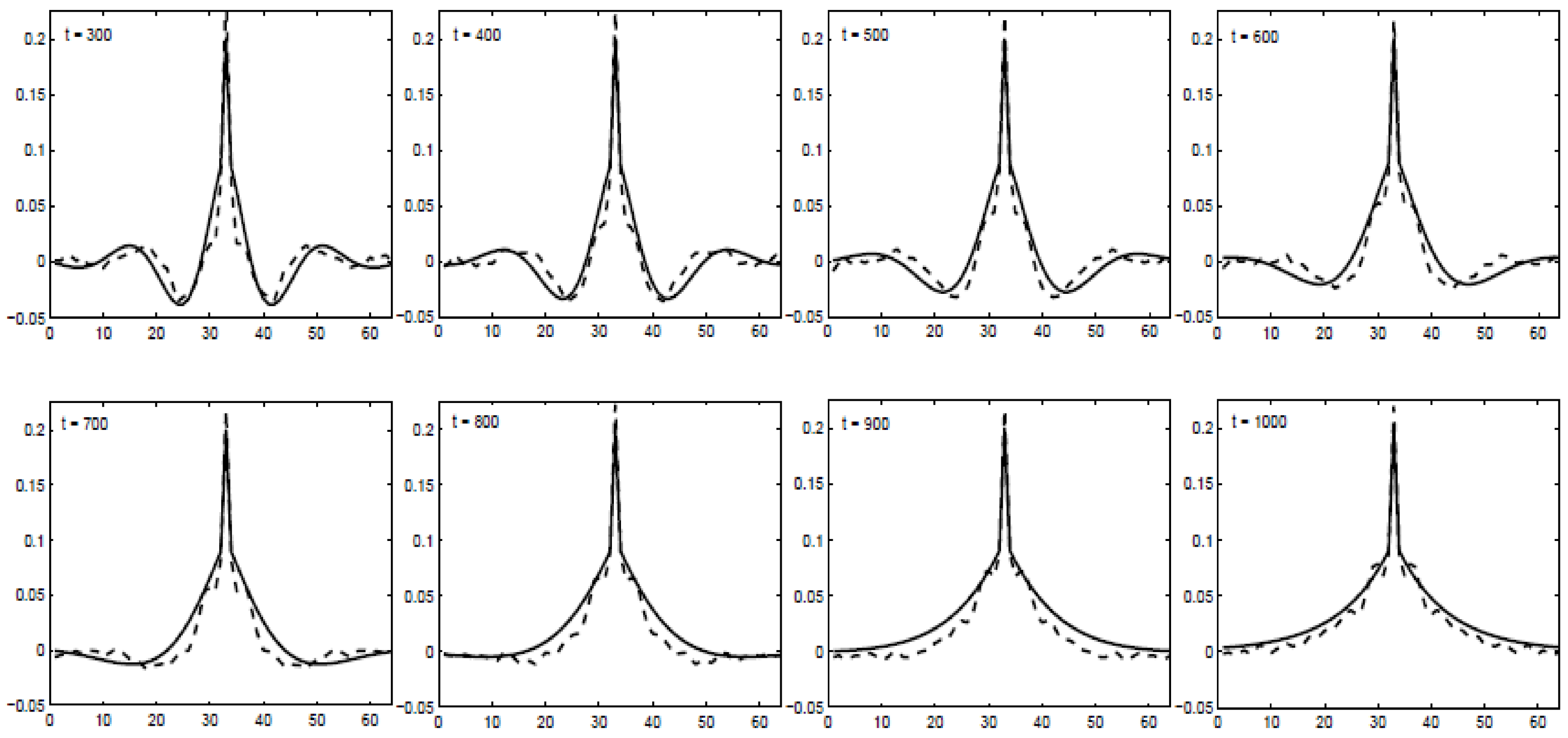

Figure 2: Rows of the true (solid) and estimated (dashed) observation error covariance matrices for the KS model with 64 observations available every 40 time steps. Background, diagonal and correlated error variances set to 0.1. 1000 ensemble members used. $N^s = 250$

Similar results are found when: Fewer ensemble members are used, observations are less frequent, error variances and ratios differ and the correlation structure varies more rapidly in time.

However, there are a number of limitations with the ETKFR method. A large number of samples may be required and the larger the number of samples the less time dependent the estimate of $\boldsymbol{R}$.

Conclusions

We have developed the ETKFR, a new method for diagnosing and incorporating spatially correlated and time-dependent observation errors in an ensemble data assimilation system. Using this method:

- We can recover the true observation error covariance in cases where the initial estimate of the matrix $\boldsymbol{R}$ is incorrect.
- We can estimate a slowly time varying observation error covariance matrix.
- The estimation of covariance structure not sensitive to the magnitude or ratio of background and observation errors.
- The inclusion of the correlated error in the assimilation reduces the analysis RMSE.

References

1. J. A.Waller, S. L. Dance, A. Lawless, N. K. Nichols, and J. R. Eyre. Representativity error for temperature and humidity using the Met Office UKV model. QJ RMetS, 2013. DOI: 10.1002/qj.2207.
2. 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, 2013. DOI: 10.3402/tellusa.v65i0.19546
3. G. Desroziers, L. Berre, B. Chapnik, and P. Poli. Diagnosis of observation, background and analysis-error statistics in observation space. QJ RMetS, 131:3385–3396, 2005.
4. J. A. Waller. Using observations at different spatial scales in data assimilation for environmental prediction. PhD thesis, University of Reading, 2013.

Acknowledgements

This work was funded by NERC as part of the National Centre for Earth Observation, the Met Office through a CASE studentship and the European Space Agency.

Contact information

- Department of Meteorology, University of Reading, Whiteknights, RG6 6AH
- Email: j.a.waller@reading.ac.uk