Estimation of observation error covariances using an ensemble transform Kalman filter

Joanne Waller

University of Reading

Supervisors: Sarah Dance, Amos Lawless, Nancy Nichols and John Eyre (Met Office)

Outline

1. Motivation and aims
2. The ETKF with R estimation
3. Results
   - frequent observations
   - infrequent observations
   - Time dependent R
4. Conclusions

Errors of Representivity

A Definition,

"Errors that arise when the observations can resolve spatial scales that the model cannot".

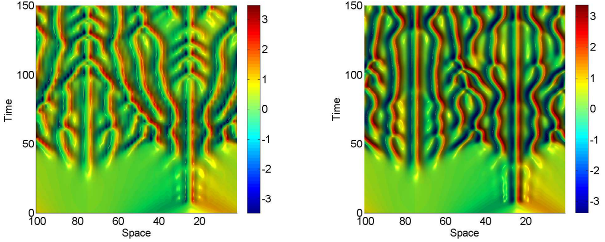
Representativity error along with errors that are introduced by the observation operator combine to make the forward error,

$$\epsilon^{\mathbf{H}} = \mathbf{y}^{\mathbf{t}} - \mathcal{H}(\mathbf{x}). \tag{1}$$

The Kuramoto Sivashinsky Equation

The Kuramoto Sivashinsky Equation,

$$u_t = -uu_x - u_{xx} - u_{xxxx}. \tag{2}$$



(a) $N = 64$ spatial points

(b) $N = 256$ spatial points

Figure : Kuramoto Sivashinsky solutions at different resolution runs

Motivation

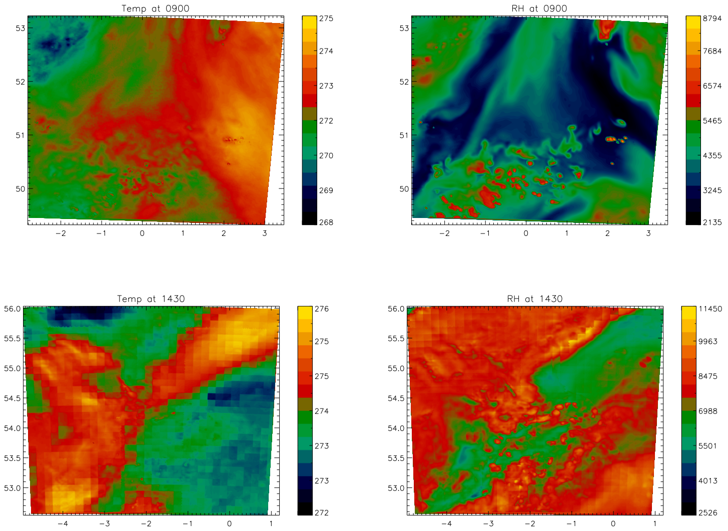


Figure : Temperature and humidity fields

Motivation

Calculating time independent representativity errors for temperature and specific humidity has shown that it is correlated [Waller et al., 2012].

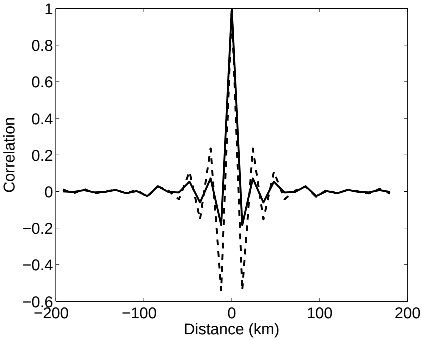


Figure : Representativity error for specific humidity using direct (solid line) and Gaussian-weighted (dashed line) observations

Motivation

It has also showed that it is state and time dependent.

Experiment Number	Observation type	Temperature RE variance ($(K)^2$)	Humidity RE variance ($(kg/kg)^2$)
1.1	Direct	4.81×10^{-3} (0.7%)	1.51×10^{-3} (1.9%)
1.3	Gaussian	8.99×10^{-4} (0.1%)	3.80×10^{-4} (0.5%)
2.1	Direct	2.21×10^{-3} (1.1%)	7.14×10^{-4} (4.0%)
2.3	Gaussian	3.96×10^{-4} (0.2%)	1.60×10^{-4} (0.9%)

Table : Representativity error (RE) variances for model with $N = 32$ all grid points observed. The values given in brackets are a comparison of the representativity error variance to the high resolution data variance.

Motivation

Representativity error variance also varies with model level height.

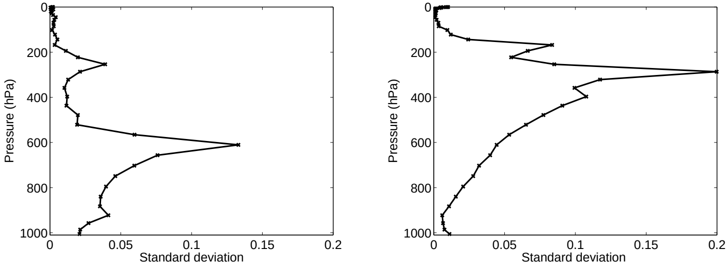


Figure : Humididty representativity error standard deviation with model level height for different synoptic situations.

Aims

- We wish to estimate time dependent R within an assimilation scheme.
- We do this by combining an ETKF with the Desroziers diagnostic.
- Subtracting the known instrument error from the estimated R gives an estimate of forward model error.

The ETKF

Initilisation

1. Determine the initial ensemble x_0^i for $i = 1 \dots N$.

Itterations

1. Forecast each ensemble member, ${x_k^f}^i = M{x_{k-1}^a}^i$.
2. Determine the ensemble mean, $\bar{x}_k^f = \frac{1}{N}\sum_{i=1}^N {x_k^f}^i$
3. Determine the ensemble perturbation matrix
   $X_k'^f = ({x_k^f}^1 - \bar{x}_k^f, \dots, {x_k^f}^N - \bar{x}_k^f)$
4. Update the ensemble mean, $\bar{x}_k^a = \bar{x}_k^f + K{d^o}_k^f$,
   where $K = X'^f(H_n{X_n'}^f)^T(H_n{X_n'}^f(H_n{X_n'}^f)^T + R_k)^{-1}$
5. Update the ensemble perturbations,
   $X_k'^a = X_k'^f(I - H_n{{X_n'}^f}^T S^{-1} H_n{X_n'}^f)^{-\frac{1}{2}}$

The Desroziers diagnostic [Desroziers et al., 2005]

The Desroziers diagnostic shows that,

$$\begin{aligned} E[d^a {d^b}^T] &= R(HBH^T + R)^{-1} E[d^b {d^b}^T], \\ &= R(HBH^T + R)^{-1}(HBH^T + R), \\ &= R, \end{aligned} \tag{3}$$

where $d_o^a = y - \mathcal{H}x^a$, and $d_o^b = y - \mathcal{H}x^b$.

This is valid if B and R used to calculate the x^a are exact. Although a reasonable estimate can be obtained even if the R and B not correctly specified.

To use in DA the estimated R must be made symmetric using $R_{sym} = 0.5(R + R^T)$.

The ETKF with R estimation

Initilisation

1. Determine the initial ensemble x_0^i for $i = 1 \dots N$.
2. Set R_0 to the known instrument error R_0^I.

Itterations

1. Forecast each ensemble member, ${x_k^f}^i = M{x_{k-1}^a}^i$.
2. Determine the ensemble mean, $\bar{x}_k^f = \frac{1}{N}\sum_{i=1}^N {x_k^f}^i$
3. Determine the ensemble perturbation matrix
   $X_k'^f = ({x_k^f}^1 - \bar{x}_k^f, \dots, {x_k^f}^N - \bar{x}_k^f)$
4. Calculate the background innovations, ${d^o}_k^f = y_k - \mathcal{H}\bar{x}_k^f$
5. Update the ensemble mean, $\bar{x}_k^a = \bar{x}_k^f + K{d^o}_k^f$,
   where $K = X'^f(H_n X_n'^f)^T(H_n X_n'^f(H_n X_n'^f)^T + R_k)^{-1}$
6. Update the ensemble perturbations,
   $X_k'^a = X_k'^f(I - H_n {X_n'^f}^T S^{-1} H_n X_n'^f)^{-\frac{1}{2}}$
7. Calculate the analysis innovations, ${d^o}_k^a = y_k - \mathcal{H}\bar{x}_k^a$
8. If $k > N^s$ update R using $R_{k+1} = \frac{1}{N^s - 1}\sum_{a=t-N^s}^{a=t} d_{o\,t}^a {d_{o\,t}^b}^T$.

Limitations

- A large number of samples may be required to calculate the observation error covariance matrix. Assuming that the correlation matrix is isotropic and homogeneous will reduce the number of samples required.
- The larger the number of samples the less time dependent the estimate becomes.
- Observations that are only available at large time intervals may represent very different synoptic situations and therefore an average over these innovations will be less related to the current synoptic situation.

1. Motivation and aims

2. The ETKF with R estimation

3. Results
   - frequent observations
   - infrequent observations
   - Time dependent R

4. Conclusions

Experiment design

- Twin experiments run with KS equation where $N = 256$ and $\Delta t = 0.25$.
- Model run the same as truth but with perturbed initial condition.
- Run with 1000 ensemble members to minimise the risk of ensemble collapse and to obtain an accurate background error covariance matrix.
- We use direct observations with added instrument error. An artificial representativity error is also added.
- Observation frequency varies between experiments, with observations available every 40 or 100 time steps.

Experiment List

- 1 - Static R, Frequent observations, Assumed $R =$ correct R
- 2 - Static R, Frequent observations, Assumed $R = R^I$
- 3 - Static R, Frequent observations, Assumed $R = \alpha R^I$
- 4 - Static R, Frequent observations, Assumed $R =$ estimated R
- 5 - Static R, Infrequent observations, Assumed $R =$ correct R
- 6 - Static R, Infrequent observations, Assumed $R = R^I$
- 7 - Static R, Infrequent observations, Assumed $R = \alpha R^I$
- 8 - Static R, Infrequent observations, Assumed $R =$ estimated R
- 9 - Slowly time varying R, Frequent observations, Assumed $R =$ estimated R

Static R, Frequent observations, Assumed R = correct R

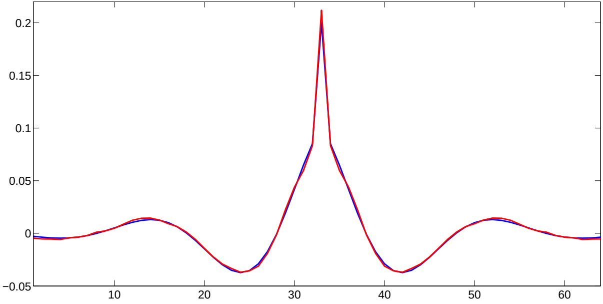


Figure : Rows of the true (blue) and estimated (red) correlation matrices for Experiment 1

Static R, Frequent observations, Assumed $R = R^I$

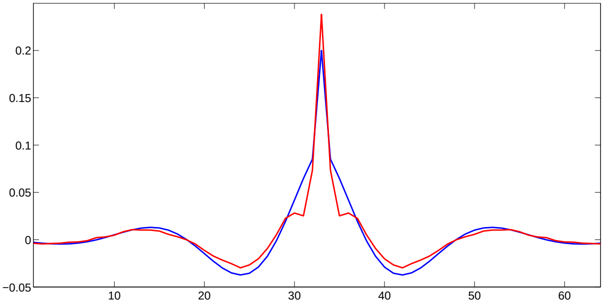


Figure : Rows of the true (blue) and estimated (red) correlation matrices for Experiment 2

Static R, Frequent observations, Assumed R $= \alpha R^l$

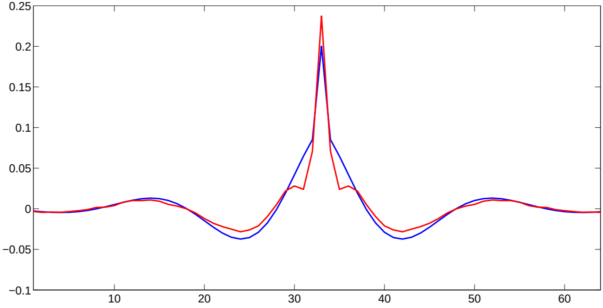


Figure : Rows of the true (blue) and estimated (red) correlation matrices for Experiment 3

Static R, Frequent observations, Assumed R = Estimated

We now use our new estimation method when R is estimated using $N^s = 250$ samples.

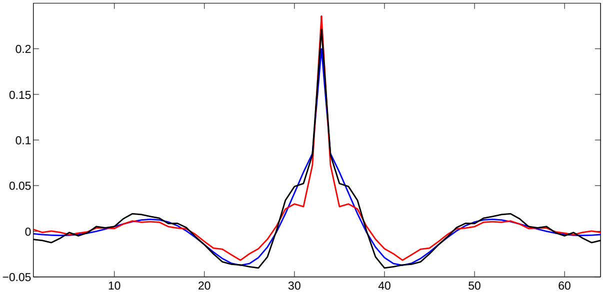


Figure : Rows of the true (blue) and estimated (red, initial estimation, black final estimation)correlation matrices for Experiment 4

Static R, infrequent observations, Assumed R = correct R

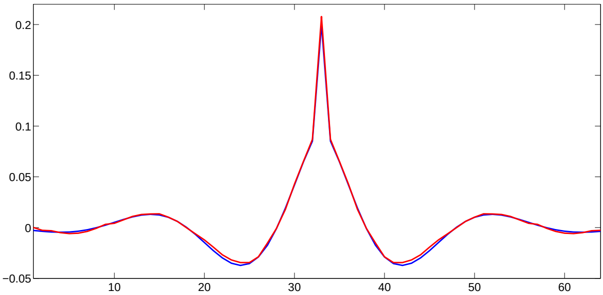


Figure : Rows of the true (blue) and estimated (red) correlation matrices for Experiment 5

Static R, infrequent observations, Assumed $R = R^I$

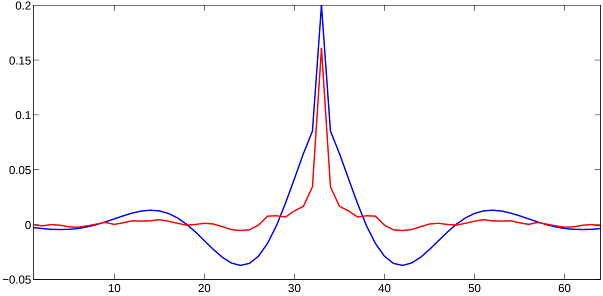


Figure : Rows of the true (blue) and estimated (red) correlation matrices for Experiment 6

Static R, infrequent observations, Assumed $\mathrm{R} = \alpha R^{I}$

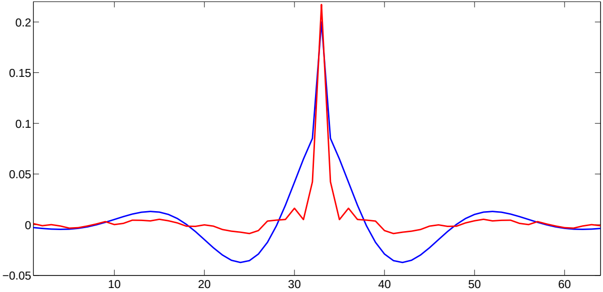


Figure : Rows of the true (blue) and estimated (red) correlation matrices for Experiment 7

Static R, infrequent observations, Assumed R = Estimated

Now use the ETKF with R estimation with $N^s = 250$.

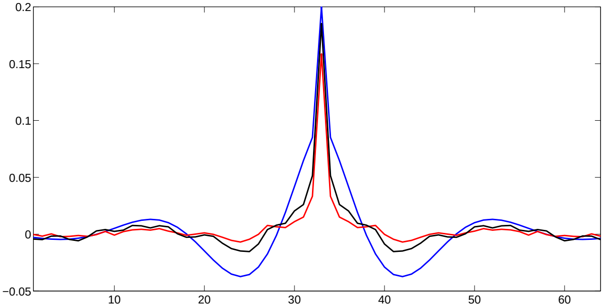


Figure : Rows of the true (blue) and estimated (red, initial estimation, black final estimation)correlation matrices for Experiment 8

Time dependent R, frequent observations

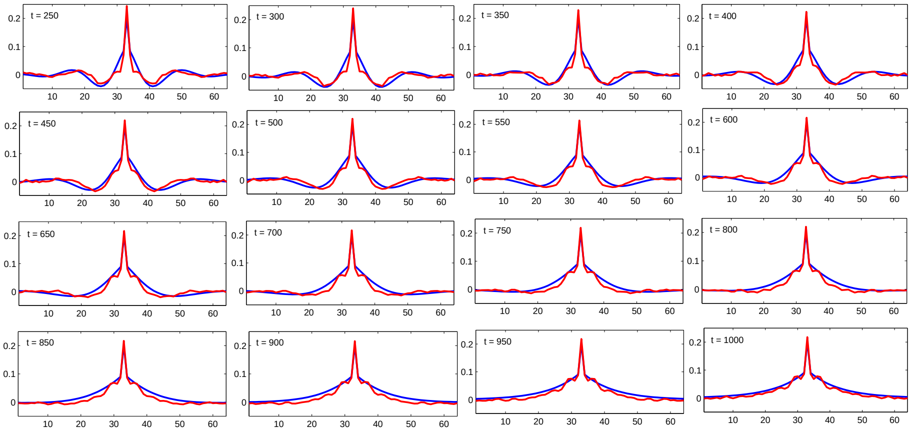


Figure : Rows of the true and estimated correlation matrices for Experiment 9

1 Motivation and aims

2 The ETKF with R estimation

3 Results
- frequent observations
- infrequent observations
- Time dependent R

4 Conclusions

Conclusions

- We have introduced an ETKF with observation error covariance matrix estimation.
- It was possible to obtain a good estimate of R using the Desroziers diagnostic. The best result was obtained when the correct matrix was used in the assimilation; however, even if the R used in the assimilation was diagonal then it was still possible to obtain a reasonable estimate of the true correlation structure.
- The method does not work as well when the observations are less frequent.
- Finally we showed that the method worked well when the true R matrix was defined to slowly vary with time.
- This suggests that the method would be suitable to give a time dependent estimate of forward or representativity error.

References

G. Desroziers, L. Berre, B. Chapnik, and P. Poli. Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society*, 131:3385 – 3396, 2005.

J. 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. Technical report, University of Reading, 2012. Department of Mathematics and Statistics Preprint, www.reading.ac.uk/maths-and-stats/research/maths-preprints.aspx.

Any questions?